Article:
Hartman, T., Kennedy, H., Steedman, R. et al. (2020) Public perceptions of good data management: findings from a UK- based survey. Big Data and Society, 7 (1). ISSN: 2053-9517

https://doi.org/10.1177/2053951720935616

Original Research Article

BIG DATA & SOCIETY

Big Data & Society
January–June: 1–16

DOI: 10.1177/2053951720935616
journals.sagepub.com/home/bds
SAGE

Public perceptions of good data management: Findings from a UK-based survey

Todd Hartman[1], Helen Kennedy[2], Robin Steedman[3] and Rhianne Jones[4]

Abstract

Low levels of public trust in data practices have led to growing calls for changes to data-driven systems, and in the EU, the General Data Protection Regulation provides a legal motivation for such changes. Data management is a vital component of data-driven systems, but what constitutes 'good' data management is not straightforward. Academic attention is turning to the question of what 'good data' might look like more generally, but public views are absent from these debates. This paper addresses this gap, reporting on a survey of the public on their views of data management approaches, undertaken by the authors and administered in the UK, where departure from the EU makes future data legislation uncertain. The survey found that respondents dislike the current approach in which commercial organizations control their personal data and prefer approaches that give them control over their data, that include oversight from regulatory bodies or that enable them to opt out of data gathering. Variations of data trusts – that is, structures that provide independent stewardship of data – were also preferable to the current approach, but not as widely preferred as control, oversight and opt out options. These features therefore constitute 'good data management' for survey respondents. These findings align only in part with principles of good data identified by policy experts and researchers. Our findings nuance understandings of good data as a concept and of good data management as a practice and point to where further research and policy action are needed.

Keywords

Good data, public perceptions, data management, data trust, personal data store, conjoint experiment

Introduction

Throughout the world, low levels of public trust in data practices have recently been identified (Edelman, 2018; Open Data Institute (ODI), 2018). There is a 'data trust deficit', it has been claimed (Royal Statistical Society (RSS), 2014), characterized by mounting concern about the potential negative consequences of the widespread use of data-driven platforms and services. Awareness of limited public trust in data practices (that is, organizational data collection, analysis and sharing and the uses to which the outcomes of these processes are put) brought about in part by high profile global failures to protect people's personal data from misuse (Cadwalladr and Graham-Harrison, 2018), has led to growing calls for changes to current data-driven systems and the structures that enable them (for example by Doteveryone, 2019a, 2019b).

The General Data Protection Regulation (GDPR), which came into effect in 2018, provides a legal motivation to improve data practices in EU countries adopting this legislation. Under GDPR, individuals have rights with regard to access and portability of

[1]Sheffield Methods Institute, University of Sheffield, Sheffield, UK
[2]Sociological Studies,University of Sheffield, Sheffield, UK
[3]Copenhagen Business School, Frederiksberg, Denmark
[4]BBC R&D, Greater Manchester, UK

Corresponding author:
Helen Kennedy, University of Sheffield, Elmfield, Northumberland Road, Sheffield S10 2TU, UK.
Email: h.kennedy@sheffield.ac.uk

their personal data. Coupled with concern about the data trust deficit, this new legislation has led to growing experimentation with alternative approaches to the management of personal data, which some believe would be better for people and society (Hall and Pesenti, 2017; O'Hara, 2019). These include personal data stores (PDSs), in which individuals personally store and manage their data, and data trusts, defined by the ODI (2019a) as 'a legal structure that provides independent stewardship of data' for the benefit of all parties.

This context has led a range of policy stakeholders to advocate for responsible and ethical data developments. In the UK, where our research took place, advocates include government centres (such as the new Centre for Data Ethics and Innovation (CDEI)), think tanks (for example Doteveryone) and independent research and advocacy organizations (such as the Ada Lovelace Institute (Ada) and the ODI). In academic circles, attention is turning to what good, responsible and ethical data might look like (for example, Daly et al., 2019). Good data management approaches, like PDSs and data trusts, are a vital part of a responsible and ethical data ecosystem, but what constitutes good data management – that is, data storage, stewardship and decision-making about sharing – is not straightforward.

Policy stakeholders in the UK, like the CDEI and Ada, claim that understanding public views about data practices is essential to ensure that data works 'for people and society' (Ada's mission) and is 'a force for good' (a CDEI aim). This also applies to data management: in order to determine what constitutes good data management, public views must be taken into account. Research into public views on good data management is therefore needed, so that these can be factored in to future data policy and practice. Yet to date, public attitudes to data management have rarely been examined, and when they have, research has focused narrowly on user feedback on specific models under development or on fictional scenarios (Sailaja et al., 2019). For this reason, our paper focuses specifically on data management, as opposed to other data practices such as data generation, collection, analysis and sharing.

The paper reports on a survey on public views on different approaches to managing personal data (that is, data related to an identified or identifiable person (GDPR, 2018)), which aimed to fill the gap identified above. The survey was administered in the UK in May 2019 to over 2000 adults. Although the GDPR was adopted in UK law after coming into force, the UK's withdrawal from the EU is causing uncertainty about future data legislation in the UK. As the UK decides what its post-Brexit data laws will look like, understanding public perceptions of good data management in the UK is extremely timely. For this reason, our survey focused on the UK. In the survey, we found that respondents dislike approaches which give commercial organizations control of personal data in return for the digital services they provide. Respondents expressed a preference for approaches that give them control over data about them, that include oversight from regulatory bodies or that enable them to opt out of data gathering altogether. These approaches are not mutually exclusive, but we separated them out for the purpose of our analysis and comment on their relationship in our conclusion. Variations of data trusts (described in detail below) were also preferable to the status quo, but not as widely preferred as approaches involving personal control, regulatory oversight or the ability to opt out. Thus personal control, oversight and the ability to opt out constituted 'good data management' for respondents in our survey.

The paper proceeds to situate our research in the context of debates about 'good data' and alternative data management approaches. We then describe our methods and discuss our findings. We conclude with reflections on the significance of our findings for conceptualizing good data and for better data management policy and practice.

Good data and alternative approaches to data management

Good data

The emerging field of critical data studies has done a good job of making visible the many troubling consequences of datafication, including increased surveillance, threats to privacy, new forms of algorithmic control, and the expansion of new and old inequalities and forms of discrimination (Iliadis and Russo, 2016; Kennedy, 2018). More recently, and against this critical backdrop, scholars have begun to consider what 'good data' alternatives might look like. One example is Daly et al.'s (2019) edited collection, *Good Data*, which was motivated by a recognition that although scholars had extensively critiqued problematic data practices, they had not considered more positive alternatives. Devitt et al. (2019) describe *Good Data* as aiming to open up 'a multifaceted conversation on the kinds of futures we want to see' and presenting 'concrete steps on how we can start realizing good data in practice'. They suggest that asking what constitutes good data is an essential step in advancing the critical scholarship which has exposed the harms and injustices that result from widespread 'bad' data practices.

Of course, 'good' is a complex concept: it could mean fair, ethical or just, or it could have other meanings, some of which acknowledge the power inequalities that shape datafication more readily than others. On the one hand, in the context of datafication, the concept of good has been used by initiatives which might be seen to depoliticize 'data relations' (Kennedy, 2016), such as charitable projects like Data For Good and DataKind. On the other hand, it is open to experienced-based interpretation, something that Andrew Sayer (2011) says is important for understanding 'why things matter to people'.

The term 'data' is not straightforward either. Daly et al. (2019) use data as a proxy for the whole DIKW model – that is, the hierarchical pyramid which has data at its base, information above that, knowledge above that and wisdom at the top. In *Good Data* and elsewhere, including in this paper, data is used as a proxy for the whole data ecology, incorporating structures, data management models, uses and consequences. As such, good data is a metaphor that extends beyond the 'high quality evidence' meaning of the term that might be more commonly used amongst data scientists and statisticians.

Good Data's editors propose a set of principles for good data practices (Devitt et al., 2019), a number of which are relevant to our focus on public perceptions of approaches to data management. Some principles highlight the importance of individual control over what happens to personal data – for example, 'data subjects must mediate data uses' and 'users must be able to understand and control their personal data'. Other principles emphasize collective needs, such as 'communal data sharing assists community participation', 'access to data promotes sustainable communal living' and 'open data enables citizen activism and empowerment'. These principles form the foundations for some of the alternative approaches to data management that we discuss below and explored in our survey.

For Daly et al., the motivation to think about good data comes from a belief that data is political and data practices should be evaluated according to whether they are used to enhance social well-being, especially for disadvantaged groups. *Good Data* thus advocates 'data methods to dismantle existing power structures through the empowerment of communities and citizens' (Devitt et al., 2019). We follow Daly et al.'s argument that good data should enhance well-being, especially amongst disadvantaged groups, because it acknowledges the politics of datafication. In order to understand whether particular approaches to data management enhance well-being, we further argue that the views of those impacted by these approaches must be considered, yet neither public perceptions nor data management feature centrally in existing debate about good data. Understanding 'bottom up' perceptions of what constitutes good data management is needed, just as it is in relation to other aspects of datafication (Couldry and Powell, 2014). We aimed to fill these gaps with the research we discuss in this paper. Furthermore, as research has shown that inequalities influence perceptions of datafication (Kennedy et al., 2020), the views of diverse populations on what constitutes a good data management model need to be examined.

Approaches to data management

In debates about data management, a number of approaches have been put forward as good alternatives to current arrangements. One is the data trust, 'a legal structure that provides independent stewardship of data' for the benefit of all parties (ODI, 2019a; see also Hall and Pesenti, 2017). According to the ODI (2019a), the trustees of a data trust 'take on responsibility to make decisions about what data to share and with whom' in order to support the trust's intended purposes and benefits. One focus of debate has been on the legal status of data trusts, as seen in the definition cited here. In the UK context, a trust is a particular legal structure which does not exist in the same form across all international jurisdictions. However, O'Hara argues that a data trust cannot be a trust in a legal sense. Rather, 'it takes inspiration from the notion of a legal trust' (O'Hara, 2019: 4). Our focus in this paper, therefore, is not on the legal dimensions of a data trust, but on its approach to stewardship. A data trust can take many forms, and data management approaches can combine features of a data trust with other features (ODI, 2019a). Table 1 below compares data trusts with other data stewardship models. A further difference relates to the type of data to be managed: some approaches are more appropriate for personal data (such as the PDS), others for open or public interest data (such as the data commons).

There are also similarities across different data management approaches. Trusts, co-operatives and commons-based approaches all involve trusted parties overseeing, managing and stewarding data on behalf of individuals and communities. In this sense (rather than in a legal sense), they are all 'trust-like'. For this reason, in our research, we explored all three of these models: (1) the data co-operative, which manages the collection and storage of its members' data, is accountable to its members and is governed by a board of representatives constituted by its members; (2) the data commons, similarly collectively motivated, which enables online access to community data which can be used for various purposes and for the benefit of all (see the decode project for an example, https://decodeproject.eu/); and (3)

Table 1. Distinguishing features of data stewardship approaches according to ODI (2019a).

Approach	Distinguishing feature
Data trusts	Takes what has been learned from the use of legal trusts. Trustees of a data trust will take on responsibility (with some liabilities) to steward data for an agreed purpose.
Data cooperatives	Takes what has been learned from cooperatives. A mutual organization owned and democratically controlled by members, who delegate control over data about them.
Data commons	Takes what has been learned from managing common pool resources – such as forests and fisheries – and applies the principles to data.
Personal data stores	Stores data provided by a single individual on their behalf and provides access to that data to third-parties when directed to by the individual.

Note: The final row of the original table, Research partnerships, has been removed because it is not relevant to our focus here.
Source: reproduced with permission from ODI (2019a).

data trusts. We differentiated between two types of trust, building on experimentation that was under way at the time of our survey (ODI, 2019a): (a) a trust governed by an independent responsible party, which makes decisions on behalf of data subjects about who accesses data, what they can do with it and under what circumstances, and (b) a trust governed by multiple independent responsible organizations which manage different types of data in different contexts (for example, one for health data, one for finance data and so on) and represent the interests of all parties involved. We consider these four models as 'trust-like' in our discussion below (see Table 2 for a full list of the data management approaches that we explored in our survey).

As can be seen in Table 2, we also explored other solutions to the perceived data trust deficit in our survey. One is the PDS, also included in Table 1. The PDS is seen as a more trustworthy approach to managing personal data than current models (for example by Janssen et al., 2019), because it enables individuals to control the processing of, access to and transfer of their personal data. Personal control has been found to be important in UK research about public attitudes to data practices: 94% of participants in a Digital Catapult (2015) survey said they wanted more control over their data. The PDS has therefore received significant attention and financial investment in recent years: notable examples include Solid (https://solid.inrupt.com/) led by Tim Berners-Lee, Databox in the UK (https://www.databoxproject.uk/) and services such as digi.me (https://digi.me/). Advocates such as the international MyData movement believe that PDSs 'empower individuals by improving their right to self-determination regarding their personal data' and that with the PDS, 'the sharing of personal data is based on trust' (MyData, nd). In contrast, critics argue that the PDS represents an individualized solution. For example, Lehtiniemi and Ruckenstein state that 'the MyData vision relies on the ethical principle of "human self-determination", treating the individual as an autonomous subject with inalienable rights and liberties' (Lehtiniemi and Ruckenstein, 2019: 6; see also Sharon and Lucivero, 2019).

Other more familiar approaches to data management also exist. Under the prevailing 'notice and consent' approach (described in our survey as the 'digital service approach' or 'status quo'), the service provider is responsible for managing personal data with users consent. Under GDPR, data controllers are mandated to notify users about the collection of their personal information and associated data practices and obtain agreement in advance. This takes the form of a privacy notice which users must consent to before they can use a service. In some instances, controls may be integrated into privacy notices allowing users to opt in or out of certain data collection practices, but it is often difficult for people to negotiate terms of use, see the extent of data practices or easily change or revoke consent. The shortcomings of the privacy notice system have been well documented (for example by Cate, 2010; Cranor, 2012; Nissenbaum, 2009; Warner and Sloan, 2013). Few people read notices in full (Obar and Oeldorf-Hirsch, 2020) and when they do, they often find them difficult to comprehend. This undermines the premise of informed consent on which the legitimacy of this approach to data management relies (Bakos et al., 2014; Nissenbaum, 2009). This approach has been described as exploitative in light of asymmetries between organizations and end users (Edwards and Veale, 2017; Zuboff, 2018) in which people have little choice but to consent to the data collection practices of digital services, if they want to participate in digital society. Despite these criticisms, this approach to data management is widely adopted across the global digital economy.

Another way to address perceived data management deficits is through regulation. Current EU and UK regulatory frameworks for data have been characterized as contradictory and unclear in a dynamic policy

Table 2. Data management approaches as described to respondents.

Name	Description
Personal data store	You are given a secure place to collect, store and manage the data about you which has been collected by other services. This is called a **personal data store, or PDS**. You have access to this data, and you can decide who else can access this data, how they can use it and under what circumstances. **The purpose of the PDS is to give you personal control over your data, which you can manage in a secure way.**
Responsible independent party	You are given a way to nominate **a responsible independent party** to oversee collection, storage and access of your personal data. They have legal responsibilities to look after your data. In line with your wishes, the nominated party can make decisions on your behalf about who accesses your data, what they can do with it and under what circumstances. **You have a say over what happens to your data, but you are not personally responsible for looking after it.**
Responsible independent organizations	**Responsible independent organizations** manage your data in different contexts (e.g. one for health data, one for finance data, etc.). These organizations make decisions about who can access your data, what they can do with it and under what circumstances. **They have legal responsibilities to manage access to your data in ways that represent the interests of *all parties involved*.**
Digital service (status quo)	You sign up to a new **digital service** (e.g. an online shop) that collects and uses your data. You are asked to agree to terms of use and a privacy policy beforehand. These describe how the service will collect, store and manage data about you. You are given settings you can alter, but you are not able to change or negotiate these terms or see how your data is used. **This approach gives services control over your data (this is what usually happens now).**
Data co-operative	You become a member of a **data co-operative** that manages the collection and storage of its members' data and is accountable to its members. As a member, you can put yourself forward to sit on a board of representatives and make decisions about who has access to members' data, how it is used and under what circumstances. Or you can vote for other co-operative members to do these things. **The purpose of the data co-operative is that your data is managed collectively, by the people whose data is in the co-operative.**
Public data commons	You access data online about your area and community using an open data platform that is accessible to all citizens under commons law. This is called a **public data commons**. The data commons collects, stores and manages access to open data which can be used for various purposes. Everyone can access and use this data, in line with the commons' rules of engagement. **The purpose of the public data commons is to make data accessible so everyone can benefit from it.**
Regulatory public body	You have been given the details of a new regulatory public body that oversees how organizations access and use data, acting on behalf of UK citizens. This public body provides oversight over how organizations collect, store and use personal data. It can hold organizations accountable for misuse (e.g. fine organizations when they breach terms of use). **The purpose of the regulatory body is to ensure that personal data are collected, stored and used in legal and fair ways.**
Data ID card (opt out)	You have the ability to choose whether to opt out of online data collection, storage and use – this is called managing your data preferences. Your data preferences are stored on a **data ID card**. You can use this card to log onto online sites. The card automatically opts you out of data collection, storage and use according to your preferences and whenever this is possible. **The purpose of the data ID card is to give people the option of opting out of having their data collected.**

environment (Hinz and Brand, nd). Previous research in the UK has found public support for better regulation of data management, such as a 2014 RSS survey which found 'more support for the government preventing misuse of personal data than an appetite to have personal control over this' (RSS, 2014: 3). GDPR has strengthened data protection regulation across EU countries that adopt it, but how the regulation is to be implemented is not entirely clear; L'Hoiry and Norris (2015) have found that data protection regulation does not easily translate from the 'law in theory' into the 'law in practice' (Galetta et al., 2016). Furthermore, as noted above, although EU laws on data protection apply to the UK during the Brexit transition period, post-Brexit data legislation in the UK is far from clear at the time of writing.

Enabling the possibility of opting in to or opting out of data collection represents another approach to data management. The PDS and variations of the data trust model enable opting in through different means, whereas opting out enables people to enact a desire not to have their data collected (Brunton and Nissenbaum, 2011). It is worth noting that although widespread adoption of either opting out or data trust models is unlikely in the current context of surveillance capitalism dominated by transnational corporations (Zuboff, 2018), these approaches play an important role in debate about future good data arrangements. For this reason, we included them in our survey. Furthermore, as noted above, approaches such as notice and consent, oversight by a regulatory body and opting out are not mutually exclusive. We separated them out in our survey to enable us to evaluate public views of them as components of good data management, and we return to a discussion of their relationship in our conclusion.

Understanding public perceptions of all of the approaches to data management discussed here is important, in order to address the data trust deficit and develop good future data practices. To date, there has been no independent and comparative research on this topic. As an active advocate for data trusts, the ODI carried out three short pilots, concluding that there is 'huge appetite' for data trusts within the organizations involved in the pilot (ODI, 2019b). The question of what members of the public, whose data is often at stake in such arrangements, think of alternative data management approaches, including data trusts, remains unanswered. In our research, we asked 'what do members of the public think constitutes good data management?' Research cited above suggested that we may find a preference for approaches premised on greater personal control (Digital Catapult, 2015), regulatory oversight (RSS, 2014) or 'trust-like' approaches (ODI, 2019b). We put the approaches discussed above to respondents in our survey to elicit their views. In the next sections, we describe our methods and findings.

Table 3. Respondent demographics compared to British Election Survey (%).

	This sample: Qualtrics May 2019	Comparison: British Election Study March 2019
Gender		
Male	47.40	45.92
Female	52.19	54.08
Other (non-binary)	0.41	–
Age		
18–34	32.64	17.00
35–54	38.13	33.68
55 or older	29.23	49.58
Education		
No formal qualification	5.17	6.44
Technical or other qualification	18.74	22.42
GCSE/A-Level (or equivalent)	48.92	40.45
University degree (or higher)	27.18	30.41
Employment status		
Full time	44.20	39.40
Part time	16.91	15.25
Not working	24.34	16.17
Retired	14.55	5.85
Household income		
$<$ £15,000	21.20	14.06
£15,000 to $<$ £30,000	32.88	31.52
£30,000 to $<$ £50,000	26.16	27.74
$>$ £50,000	19.76	22.03
Ethnicity		
White	90.63	95.74
BAME	9.37	4.26
Disability		
Disabled	20.94	31.35
Non-disabled	79.06	68.65
Total %	100.00	100.00
N	2169	30,842

Note: Our data was collected from members of a self-selected Internet panel by Qualtrics in May 2019. British Election Study (BES) data was collected by YouGov in March 2019. Respondents who provided a 'don't know' answer or refused to answer a question are not included in these totals. Not all percentages sum to 100 due to rounding.

Respondents' existing knowledge and views about data practices

In May 2019, 2169 respondents living within the UK completed our online survey. The survey focused on what participants thought about the eight approaches to managing data listed in Table 2. We collected data from diverse respondents from across the UK (for a full demographic breakdown, see Table 3). Respondents were recruited by Qualtrics using opt-in methods, the sample demographics of which compare favourably with other reputable Internet panels such as the British Election Study conducted by YouGov (see column 2 in Table 3). Qualtrics partners with online sample providers to recruit diverse respondents for research purposes. Researchers have found that Qualtrics approximates probability-based samples reasonably well in terms of demographic characteristics and responses to other socio-political questions (Zack et al., 2019). It should be noted that surveys conducted online using an Internet panel like Qualtrics are likely

to recruit respondents who are capable technology users. This was confirmed in answers to related questions: 94.6% indicated that they were confident using devices to do things online, 98.9% stated they used the Internet daily and only 8.5% of respondents indicated that they were not users of at least one of the major social media platforms.

Before rating the approaches, respondents completed knowledge questions to gauge their familiarity with and understanding of concepts relevant to the survey. We presented participants with a series of statements about personal data, open data and the GDPR and asked them to identify whether each statement was true or false. These statements were used to assess their knowledge about relevant issues and evaluate responses to later questions in light of these responses. Some of these statements were reverse worded to account for potential agreement bias. Respondents appeared most knowledgeable about the concept of personal data, with the vast majority correctly answering questions related to its definition: more than 7 out of 10 respondents answered these questions correctly. Respondents were least knowledgeable about open data: less than half were able to correctly answer two questions on this topic. Results were mixed concerning familiarity with and understanding of GDPR: 93% of the sample correctly answered a question about its main purpose and 53% provided correct answers to a question about data portability (see Table 4).

Once completed, we provided respondents with the answers to these questions to ensure that everyone began subsequent sections with the same general information about the topic. We also included questions on attitudes towards how personal data is collected, stored, used and shared by organizations, to gauge respondents' views on a broad range of related issues and enable us to analyse whether attitudes were indicators of preference. We asked participants to indicate on a five-point Likert scale whether they agreed or disagreed with a series of statements. Respondents were concerned about the privacy (84.6% agreement) and security (84.2%) of their personal data. They wanted to be able to exercise their rights (92.1%) and have more control over their data (89.0%). In particular, they were concerned about how their personal data is used by organizations (86.9%), and they wanted companies to be held accountable if it is misused (96.1%). Respondents were against commercial organizations using personal data to generate profit (78.3%). Only around half of the respondents supported sharing personal data for use in research in the public interest (52.7%). Around two in three wanted data to be used for the social good (68.8%). Most want data to be managed, analysed and gathered in ethical ways (84.0%). A full list of statements and responses can be seen in the Supplemental Appendix.

In another part of the survey, we asked participants about the types of data-driven apps and services that they would like to see developed in the future, inviting them to select services from a list or add their own. Types included related to health, well-being, the environment and education. When we asked respondents who they would like to see provide these services, most said they preferred governmental or publicly-funded organizations – 46% and 40% of respondents selected these options, compared to 18% selecting commercial organizations in a question where respondents could select as many options as they wished.

The questions discussed thus far were asked to aid our analysis. Existing research has highlighted that knowledge levels influence public views about data practices (Digital Catapult, 2015; Doteveryone, 2018) and as such establishing existing knowledge levels was necessary. Standard demographic questions were asked to enable us to explore whether different groups of people have different views about good data management. Responses to questions about future data-driven apps and services indicate what might constitute good data management for respondents: personal control; the ability to exercise one's rights; accountable, prosocial uses of data; and oversight by a public body.

Views on approaches to data management

Examining respondents' views about data management approaches was at the heart of our survey, and we used three different methods to do this. Our first method asked respondents to rate four randomly selected data management approaches (presented one at a time) using a Likert scale ranging from 0 (poor) to 10 (excellent). This method is commonly used in surveys, yet assigning a numeric value on an 11-point scale can be difficult for some respondents. To address this issue, our second method of assessing preferences used an innovative approach called a conjoint experiment (Hainmueller et al., 2014). A conjoint experiment works by presenting respondents with options randomly generated from a list. The task involves comparing items side-by-side and then choosing the preferred option. This forced choice design simplifies the decision facing respondents (Hainmueller and Hopkins, 2015; Pelzer, 2019). We used a single-attribute conjoint experiment in which participants were presented with two randomly selected approaches from the list of eight (see Table 2 for the exact wording of each model) and asked them to select the approach that they preferred from the pair. This paired selection task was repeated three times for each

Table 4. Percentage of knowledge questions answered correctly.

Question (*correct response*)	% Correct
The General Data Protection Regulation (GDPR) governs the processing of personal data (collection, storage and use). *(True)*	93.1
Any information that can be used to identify an individual is personal data. *(True)*	92.2
Location data collected by your mobile phone is not personal data. *(False)*	73.4
The General Data Protection Regulation (GDPR) does not give you the right to access the personal data organizations hold about you. *(False)*	72.2
There are still no financial penalties for companies that do not comply with the General Data Protection Regulation (GDPR). (*False)*	69.0
The General Data Protection Regulation (GDPR) allows for 'data portability' meaning that you can take your data from one organization and give it to another. *(True)*	52.6
Open data does not generally include personal data. *(True)*	48.9
Open data can only be used, modified and shared for non-commercial purposes. *(False)*	48.2

Table 5. Example of the single-attribute conjoint experiment.

Option A	Option B
You are given a secure place to collect, store and manage the data about you which has been collected by other services. This is called a personal data store, or PDS. You have access to this data, and you can decide who else can access this data, how they can use it and under what circumstances. The purpose of the PDS is to give you personal control over your data, which you can manage in a secure way.	You are given a way to nominate a responsible independent party to oversee collection, storage and access of your personal data. They have legal responsibilities to look after your data. In line with your wishes, the nominated party can make decisions on your behalf about who accesses your data, what they can do with it and under what circumstances. You have a say over what happens to your data, but you are not personally responsible for looking after it.
Based on these descriptions, which option for managing data would you prefer? ☐ Option A ☐ Option B	

respondent. Table 5 provides an example of the single-attribute conjoint experiment used in this study, which allowed us to evaluate how respondents rated the approaches in comparison to one another.

Our third and related method for assessing respondents' views of the data management approaches was to ask them to complete a multiple-attribute conjoint experiment. This differed from the second method we described above in allowing us to compare different factors that may affect the decision to select one data management approach over another. We accomplished this by randomly combining multiple factors into data management profiles to assess the relative effect of each specific factor on preferences. We asked respondents to express preferences for scenarios generated from a combination of factors identified as significant in previous research (for example Kennedy et al., 2015):

- Type of data (for example, medical, financial, media consumption);
- What management arrangements mean for the individual (for example, full control over what happens to data, knowing what data is held about them, by whom and what they do with it);
- Use and beneficiaries of the data (for example, personal insights, generate profit, benefit society).

In addition, we included who has control (for example, individual, trustee, commercial organization) as a factor, as this is relevant to our focus on data management. An example of our multiple-attribute conjoint experiment is provided in Table 6 (the full survey and stimulus materials are available in the Supplemental Appendix).

Preferences in relation to approaches

Of the eight approaches to data management that we presented to respondents, three were consistently rated highly. The most preferred approach was the PDS, described in the survey as 'a secure place to collect, store and manage the data about you which has been collected by other services' which would give individuals control over their personal data (see Table 7 for

Table 6. Example of the multiple-attribute conjoint experiment.

	Option A	Option B
In this scenario, the data is	Medical data	Financial data
The data is controlled by	You	A trustee like a city council or the government
You will be able to	Have full control over what happens to it	Know what data is held about you, by whom and what they do with it
The data will be used for these reasons, and generate these benefits	So you can get insights and value from your personal data	So an organization can use your data to benefit the public
Based on the descriptions, which of these options would you prefer?		
☐ Option A		
☐ Option B		

Table 7. Mean ratings on a scale from 0 to 10 for each data management model.

Model	Mean rating
Personal data store	7.7
Regulatory public body	7.6
Data ID card (with clear opt-out options)	7.5
Responsible independent organizations	6.4
Public data commons	6.3
Responsible independent party	6.2
Data co-operative	5.9
Status quo	4.9

mean ratings of each model). Responses to questions about views on data uses suggest that the possibility of greater individual control may be why this approach was highly rated: 86.9% of respondents agreed with the statement 'I want more control over how my personal data is used by organizations', and 89.0% agreed with the statement 'I want more control over my personal data'. As noted above, previous research by Digital Catapult (2015) also highlighted the importance of personal control.

After the PDS, the next highest rated approach involved a regulatory public body overseeing 'how organizations access and use data, acting on behalf of UK citizens' in order to 'ensure that personal data are collected, stored and used in legal and fair ways'. As noted above, elsewhere in the survey, we asked respondents who they would like to see provide new data-driven services 'for the public good' and most selected governmental organizations (46% of respondents), followed by publicly-funded organizations (40%). This reinforces the finding that oversight of data by a public regulatory body was a strong preference for our respondents.

The high rating of this model by respondents suggests a preference for legally enforceable safeguards alongside the personal control of data offered by the PDS. This finding was confirmed in responses to questions about views on data uses, in which 96.1% of respondents agreed with the statement 'I want companies to be held accountable if they misuse my personal data'. Realizing this statement requires governance, which may explain respondents' strong preference for data management to be overseen by a regulatory body. In contrast to the RSS (2014) survey cited above, which found more support for governance than personal control, we found a strong preference for both. The high ranking of both the PDS and oversight by a regulatory public body suggests that *both* personal control *and* oversight are important principles of good data management for respondents.

We described the approach that would allow people to opt out of having their data collected as a 'Data ID Card', to give material form to a means for opting out of data collection. This approach was ranked third overall. The relatively high ranking of this model reinforces the importance of individual control over data amongst our respondents. It also shows that respondents would be willing to opt out of data gathering, indicating strong dissatisfaction with current data arrangements.

We explored respondents' views on data management in multiple ways in the survey, to ensure reliability of findings. We found that the results of the single-attribute conjoint experiment corroborated the findings discussed above. This experiment asked respondents to choose the option that they preferred from a randomly generated pair of approaches, the results of which are presented in Figure 1. The plotted points provide the change in the probability of selecting an approach relative to the status quo (that is, digital services having control over people's data). The vertical dotted line indicates the digital service/status quo baseline; points to the right of the dotted line indicate an increase in the probability of choosing that particular approach relative to the baseline. The lines around each side of plotted points are 95% error bars, indicating uncertainty

Figure 1. Results from the single-attribute conjoint analysis.

around each value, which derives from the fact that our survey is based on a sample of the population.

As with the individual ratings task, this experiment revealed that the top three preferred approaches are the PDS, opting out and oversight by a regulatory public body, in that order of preference. There was at least a 30% point increase in selecting any of the top three data management approaches compared to the status quo/'notice and consent' approach. This is a significant number, both statistically and substantively. The approaches that did not offer personal control or regulatory oversight, which we describe above as 'trust-like', had lower mean scores than those that did offer such features, in both the rating task and the single-attribute conjoint experiment. These include approaches overseen by a public data commons, a data co-operative, multiple responsible independent organizations or a specific responsible independent party. Trust-like approaches were preferable to the status quo, but less preferable than those based on the concepts of personal choice, control and regulation. These approaches may have received lower ratings because they were less familiar to respondents than approaches based on the more commonplace concepts of choice, control and regulation. As noted above, in the knowledge questions with which we opened the survey, respondents demonstrated limited knowledge of open data, the principles of which influence data trust approaches. In addition, elsewhere in the survey, only 39.3% of respondents agreed with the statement 'I'm in favour of open data'. This relatively low level of support for open data could result from the low levels of understanding of open data that we also identified. Together, these findings may explain the lower mean scores for the 'trust-like' data management approaches that we presented to respondents.

It is striking that respondents preferred all other approaches to a 'digital services model' that 'gives services control over what happens to your data'. With an average rating of just 4.9 out of 10, this suggests that respondents are unsatisfied with services and organizations controlling data. Combined with the high rating

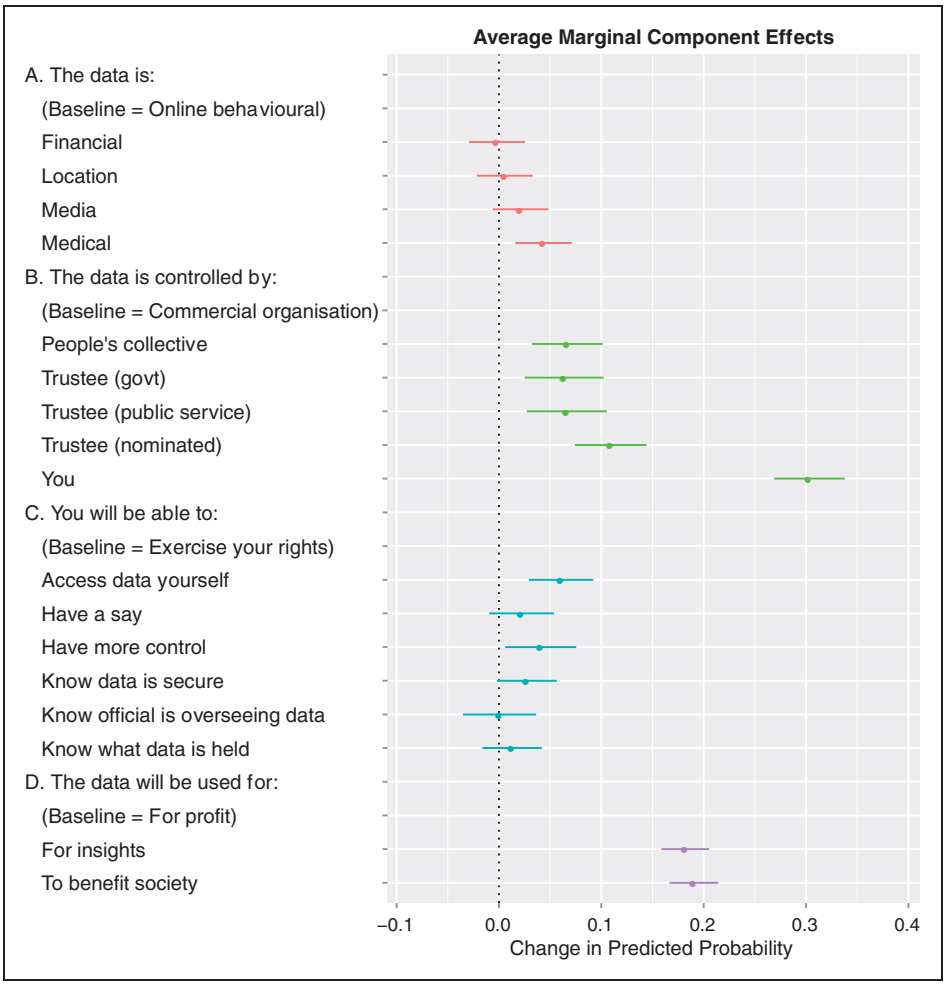

Figure 2. Results from the multiple-attribute conjoint analysis.

of the opt out model, and strong support for statements expressing concern about data management issues, these findings show that current arrangements require radical change in order to win public support.

Preferences in relation to data handling scenarios

We also used a multiple-attribute conjoint experiment, which compared the significance of a number of factors in data handling scenarios – including types of data, uses of data and related benefits (identified as significant in previous research (Kennedy et al., 2015)) and control arrangements and what these enable – to assess preferences towards data management approaches. Figure 2 displays the results from this conjoint experiment. As with the single-attribute conjoint analysis, in the figure, we present results which show the change in the probability of selecting a profile with particular characteristics relative to a baseline, this time for each of the attributes we included in the scenarios.

Figure 2 demonstrates that the most important factor influencing responses to the multiple-attribute conjoint experiment was the locus of control over data – respondents want control to rest with them. The probability of respondents selecting a data management scenario that gives them control over their own data increased by 30% points relative to the baseline (that is, a commercial organization controls the data). Thus, personal control played a key role in this experiment, just as it did in evaluations of data management approaches (as seen in Table 7) and in responses to statements about data use and management. As we discovered throughout the survey, respondents preferred scenarios in which anyone other than a commercial organization was responsible for controlling their data. In this experiment, there was little notable differentiation among the alternative controllers that we presented, apart from respondents themselves, for whom a significant preference was expressed.

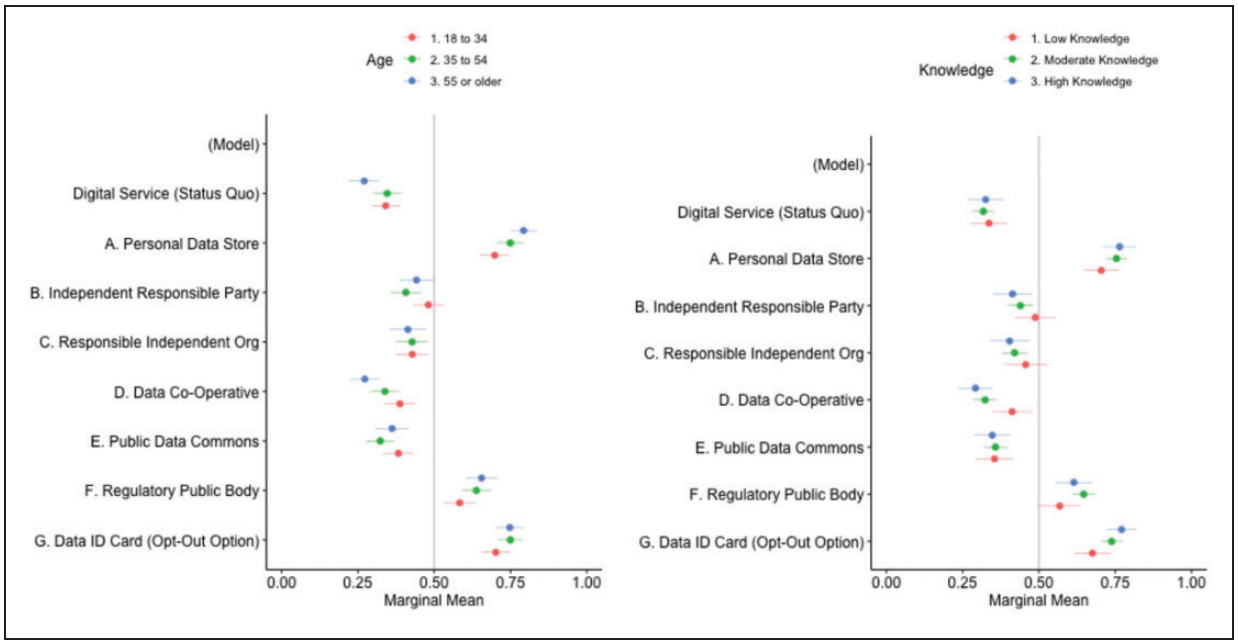

Figure 3. Subgroup responses to the single-attribute conjoint experiment by age group and existing knowledge.

The other significant factor in this experiment related to uses of data and beneficiaries. Respondents preferred scenarios in which data would be used for insights or to benefit society rather than for profit, which is consistent with findings from other surveys (e.g. Doteveryone, 2018). The effect sizes for these factors were in the medium range, with a change in the probability of selecting that profile of 0.15 or greater. In other words, there is a 15-percentage point increase in the chance that a particular profile would be selected when it provided personal insights or benefits to society compared to profit. Other factors were not as impactful, as Figure 2 demonstrates. For instance, respondents did not significantly differentiate in relation to what management arrangements mean for the individual (for example, giving them control over what happens to data or enabling them to know what data is held about them), as seen in Figure 2(c). Finally, this experiment confirmed the finding from elsewhere in the survey that respondents do not like their personal data to be controlled by commercial organizations (Figure 2 (b)) or used for profit (Figure 2(d)). As Figure 2 shows, all other scenarios were preferable to this one.

Differences amongst respondents

Recent research has demonstrated that people experience datafication differently. Ethnicity, gender, poverty and their intersections have been shown to impact people's experiences of data practices (Eubanks, 2017; Noble, 2018). There is much less research into whether social inequalities influence *perceptions* of data practices (Kennedy et al., 2020 is one exception). Because of this, we analysed whether these and other characteristics, including existing knowledge of data-related matters, had an impact on respondents' views of data management approaches. This latter variable, knowledge, was indeed a significant predictor of preferences in relation to some of the approaches (see the Supplemental Appendix for full results). In the ratings exercise, for example, knowledgeable respondents preferred approaches that offered more control and/or oversight over personal data by a regulatory public body than less knowledgeable respondents, who indicated a slightly higher preference for the status quo, which gives digital services control over their data. This effect was relatively small (about a half point difference on a 10-point scale). Age also had a significant impact on ratings of approaches: younger respondents rated the status quo model higher than those who were aged 35 years and over (about 1 point higher mean rating on a 10-point scale). Thus, differences relating to age and existing knowledge mattered, but not a great deal. Apart from these two findings, there were no other clear differences in evaluations by demographic subgroups within the sample. In other words, we did not find that gender, ethnicity, educational attainment, employment status or household income were significant predictors of preferences.

Similar subgroup differences were observed in the single-attribute conjoint experiment, presented in Figure 3 (the full set of comparisons is available in

the Supplemental Appendix). This figure plots the average proportion of respondents selecting each data management model, also known as marginal means, by age and knowledge. By design, marginal means average 0.5. In other words, if responses were simply randomly chosen, there is a 50:50 chance that a given response is selected. Values above 0.5 tell us that respondents prefer a given approach, and values below 0.5 indicate that respondents do not like the approach. A value of 0 would tell us that the approach was never selected; a value of 1 means that it was always selected. As with previous figures, Figure 3 also includes error bars.

While the plot points for various demographic subgroups were for the most part grouped closely together, indicating consistency in responses, there are some exceptions. One is age, which appears to have some influence on preference. Respondents in the 18–34 years age group were less swayed by the PDS, oversight by a regulatory public body and the opt out option than respondents aged 35 years and over, although younger respondents still preferred these approaches to the others presented to them. This is indicated in Figure 3 by the closer proximity to the 0.5 value for younger respondents. Less knowledgeable respondents, in general, were also less likely to differentiate among the approaches. Again, this is shown in the closer proximity of their marginal means to the 0.5 vertical line. The effects of both of these variables, however, are relatively small, as we observed with responses to other survey items.

Discussion and conclusions

Our research asked 'what do members of the UK public think constitutes good data management?' Our findings suggest that personal data, oversight from regulatory bodies and the choice to opt out of data gathering are the main components of good data management from the perspective of the UK public. Another important finding is that respondents dislike approaches in which commercial organizations control and profit from personal data in exchange for digital services. As noted above, these approaches to data management are not mutually exclusive. Under GDPR, the dominant 'notice and consent' model should include opt out options and oversight from regulatory bodies. In this context, we draw three conclusions from our findings.

First, our research suggests that organizations which handle personal data and policy-makers in this domain need to accept that current arrangements are not acceptable. People like the idea of choice, control and oversight, and they do not like commercial organizations controlling and profiting from their personal data. Second, given that some of preferred features are provided for under GDPR, which continues to be implemented in the UK at the time of writing, our findings raise questions for future research about the relationship between the 'law in theory' and the 'law in practice' (Galetta et al., 2016). These include questions about whether people perceive the existing arrangements as 'good' but in need of better enforcement, or whether greater oversight by regulators and more stringent regulations would be preferred.

Third, we need to think carefully about what respondents' preference for more control over their personal data might look like in practice. In previous qualitative research that we have undertaken, participants expressed concern about the burden of decision-making that a PDS approach might impose upon them as individuals (Steedman et al., 2020). Offloading the responsibility for good and informed data management decision-making onto citizens may therefore be problematic. Effective approaches to greater personal control need further research. Our research has identified what users want; further research into how to realise this in practice is needed.

A further finding from our survey is that not all alternatives to data management are rated equally by respondents. Although they preferred all alternatives to the status quo, they expressed a greater preference for some than for others. Data trust-like approaches – a public data commons, a data co-operative, oversight by a responsible independent party or organizations – were ranked below PDS, regulatory and opt out approaches. These findings were consistent across different methods used in the survey. We cannot therefore conclude that there is a 'huge appetite' for data trusts amongst the public, as the ODI suggests exists amongst organizational stakeholders, based on their pilot (ODI, 2019b). Further research is needed to explore the reasons for this, although some speculation is possible. Data trust-like approaches may have been rated lower than other approaches because they were less familiar to respondents than approaches based on the more commonplace concepts of control, opting out and regulation. Respondents' limited knowledge of and support for open data, the principles of which inform data trusts, was evidenced in answers to diverse questions in the survey. This might explain respondents' lesser preference for these approaches.

Existing knowledge and age had an impact on evaluations of approaches, but the effects of these factors were relatively small. The fact that less knowledgeable respondents were less likely to differentiate amongst approaches might suggest that with good information, more differentiation of approaches might result. But the relationship between information, understanding and perceptions of data practices is complex, and previous research has shown that information and

understanding are not necessarily the solution to the data trust deficit (Steedman et al., 2020). Here again, further research is needed to understand the relationship between knowledge about and preference for data management approaches in greater depth.

Our research indicates that public views of good data management align only in part with the principles of good data identified by experts and commentators. Devitt et al.'s (2019) principles 'users must be able to understand and control their personal data' and 'data subjects must mediate data uses' were confirmed by our respondents strong preference for a PDS model or an opt out option to give them control over what happens to their data. However, collective principles such as 'communal data sharing assists community participation', 'access to data promotes sustainable communal living', and 'open data enables citizen activism and empowerment', represented in data co-operative and public data commons approaches, were not as widely preferred, although respondents did indicate support for pro-social uses of data. Respondents' evaluations of what constitutes good data management did not align with those experts who argue that data trusts represent a model of good data either, given that the trust-like approaches that we presented to them were not the most preferred options. A major contribution of our research, then, is that it nuances understandings of good data as a concept and of good data management as a practice.

In some ways, the UK is in a unique position when it comes to data management futures, given current uncertainty about post-Brexit data regulation. This situation provides the UK government with an opportunity to heed what the public wants, which has been the main focus of our paper. We found a 'huge appetite' for alternatives to commercial control of personal data amongst our respondents, and a clear indication of what constitutes good data management for them. The UK government could choose to implement good data management approaches which have public support, but this would require investment of resources for technical development and for further public consultation. By contrast, disregard for public views about what constitutes good data management would perpetuate distrust, and this would likely have consequences both for government and for organizations that are trying to work with data in ways that are good, ethical and responsible. In many ways, these conclusions are not unique to the UK. Many countries face similar challenges relating to trust, and research on attitudes to data practices in general has found similar levels of concern across countries (for example Edelman, 2018; European Commission, 2019; ODI, 2018; PEGA, 2019). Further research is needed across the globe to explore why particular data management preferences exist, and global action is also needed, from data policy-makers and practitioners, to respond to public concerns.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by a grant from the Arts and Humanities Research Council, award number AH/S012109/1, and BBC Research and Development.

ORCID iDs

Helen Kennedy https://orcid.org/0000-0003-0273-3825
Robin Steedman https://orcid.org/0000-0003-1033-9318

Supplemental Material

Supplemental material for this article is available online.

References

Bakos Y, Marotta-Wurgler F and Trossen DR (2014) Does anyone read the fine print? Consumer attention to standard-form contracts. *The Journal of Legal Studies* 43(1): 1–35.

Brunton F and Nissenbaum H (2011) Vernacular resistance to data collection and analysis: A political theory of obfuscation. *First Monday* 16(5).

Cadwalladr C and Graham-Harrison E (2018; March 17) Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*. Available at: www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election (accessed 18 March 2018).

Cate FH (2010) The limits of notice and choice. *IEEE Security & Privacy Magazine* 8(2): 59–62.

Couldry N and Powell A (2014) Big data from the bottom up. *Big Data and Society* 1(1): 1–5.

Cranor FL (2012) Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *Journal on Telecommunications and High Technology Law* 10(2): 273–307.

Daly A, Devitt SK and Mann M (2019) *Good Data*. Amsterdam, the Netherlands: Institute of Network Cultures.

Devitt SK, Mann M and Daly A (2019) The 'Good Data' Project. Available at: www.networkcultures.org/blog/2019/01/11/principles-of-good-data/ (accessed 5 June 2020).

Digital Catapult (2015) Trust in personal data: A UK review. Report by Digital Catapult, London, UK.

Doteveryone (2018) People, power and technology: The 2018 digital attitudes report. Available at: www.understanding.doteveryone.org.uk (accessed 5 June 2020).

Doteveryone (2019a) Engaging the public with responsible technology: Four principles and three requirements. Available at: www.doteveryone.org.uk/download/3225/ (accessed 5 June 2020).

Doteveryone (2019b) Better redress: Building accountability for the digital age: An evidence review from Doteveryone. Available at: www.doteveryone.org.uk/wp-content/uploads/2019/12/Better-redress-evidence-review.pdf (accessed 5 June 2020).

Edelman (2018) Edelman Trust Barometer 2018. Available at: www.edelman.co.uk/research/edelman-trust-barometer-2018-uk-findings (accessed 5 June 2020).

Edwards L and Veale M (2017) Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for. *Duke Law and Technology Review* 16(1): 18–84.

Eubanks V (2017) *Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor*. New York, NY: St Martins Press.

European Commission (2019) Special Eurobarometer 487a. Summary – The General Data Protection Regulation. Available at: www.ec.europa.eu/commfrontoffice/publicopinionmobile/index.cfm/Survey/getSurveyDetail/surveyKy/2222 (accessed 5 June 2020).

Galetta A, Fonio C and Ceresa A (2016) Nothing is as it seems. The exercise of access rights in Italy and Belgium: Dispelling fallacies in the legal reasoning from the 'law in theory' to the 'law in Practice'. *International Data Privacy Law* 6(1): 16–27.

GDPR (2018) Personal data. Available at: www.gdpr-info.eu/issues/personal-data/ (accessed 5 June 2020).

Hainmueller J and Hopkins DJ (2015) The hidden American immigration consensus: A conjoint analysis of attitudes toward immigrants. *American Journal of Political Science* 59(3): 529–548.

Hainmueller J, Hopkins D and Yamamoto T (2014) Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis* 22(1): 1–30.

Hall W and Pesenti J (2017) *Growing the Artificial Intelligence Industry in the UK*. London, UK: DCMS.

Hinz A and Brand J (nd) Data policies: Regulatory approaches for data-driven platforms in the UK and EU. Available at: www.datajustice.files.wordpress.com/2020/01/data-policies-research-report-revised.pdf (accessed 5 June 2020).

Iliadis A and Russo F (2016) Critical data studies: An introduction. *Big Data & Society* 3(2): 1–7.

Janssen H, Cobbe J, Norval C, et al. (2019) Personal data stores and the GDPR's lawful grounds for processing personal data. *Zenodo*. Epub ahead of print 29 May 2019. DOI: 10.5281/zenodo.3234902.

Kennedy H, Elgesem D and Miguel C (2015) On fairness: User perspectives on social media data mining. *Convergence* 8(6): 859–876.

Kennedy H (2016) *Post, Mine, Repeat: Social Media Data Mining Becomes Ordinary*. Basingstoke, UK: Palgrave Macmillan.

Kennedy H (2018) Living with data: Aligning data studies and data activism through a focus on everyday experiences of Datafication'. *Krisis: Journal for Contemporary Philosophy*. Available at: www.krisis.eu/living-with-data/ (accessed 5 June 2020).

Kennedy H, Steedman R and Jones R (2020) Approaching public perceptions of datafication through the lens of inequality: A case study in public service media' *information. Communication and Society*. Epub ahead of print 4 March 2020. DOI: 10.1080/1369118X.2020.1736122.

Lehtiniemi T and Ruckenstein M (2019) The social imaginaries of data activism. *Big Data & Society* 6(1): 1–12.

L'Hoiry X and Norris C (2015) The honest data protection officer's guide to subject access requests. *International Data Privacy Law* 5(3): 190–214.

MyData (nd) Homepage. Available at: www.mydata.org/ (accessed 5 June 2020).

Nissenbaum H (2009) *Privacy in Context: Technology, Policy and the Integrity of Social Life*. Palo Alto, CA: Stanford University Press.

Noble S (2018) *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NY: New York University Press.

Obar AJ and Oeldorf-Hirsch A (2020) The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society* 23(1): 128–147.

ODI (2018) Who do we trust with personal data? Available at: www.theodi.org/article/who-do-we-trust-with-personal-data-odi-commissioned-survey-reveals-most-and-least-trusted-sectors-across-europe/ (accessed 5 June 2020).

ODI (2019a) Data trusts: Lessons from three pilots. Available at: www.docs.google.com/document/d/118RqyUAWP3WIyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit# (accessed 5 June 2020).

ODI (2019b) Huge appetite for data trusts. Available at: www.theodi.org/article/huge-appetite-for-data-trusts-according-to-new-odi-research/ (accessed 15 April 2019).

O'Hara K (2019) *Data trusts: Ethics, architecture and governance for trustworthy data stewardship*. White Paper. Available at: www.eprints.soton.ac.uk/428276/ (accessed 5 June 2020)..

PEGA (2019) GDPR: Show me the data survey reveals EU consumers poised to act on legislation. Available at: www.pega.com/system/files/resources/2019-07/GDPR-Show-Me-The-Data-eBook.pdf (accessed 15 April 2019).

Pelzer E (2019) The potential of conjoint analysis for communication research. *Communication Research Reports* 36(2): 136–147.

RSS (2014) Trust in data and attitudes toward data use/data sharing. Available at: www.statslife.org.uk/images/pdf/rss-data-trust-data-sharing-attitudes-research-note.pdf (accessed 24 February 2019).

Sailaja N, Colley J, Crabtree A, et al. (2019) The living room of the future. In: *Proceedings of TVX 2019: The ACM conference on interactive experiences for television and online video*. Salford, UK, 5 June 2019.

Sayer A (2011) *Why Things Matter to People: Social Science, Values and Ethical Life*. Cambridge, UK: Cambridge University Press.

Sharon T and Lucivero F (2019) Introduction to the special theme: The expansion of the health data ecosystem – Rethinking data ethics and governance. *Big Data & Society* 6(2): 1–5.

Steedman R, Kennedy H and Jones R (2020) Complex ecologies of trust in data practices and data-driven systems. *Information, Communication and Society*. Epub ahead of print 8 April 2020. DOI: 10.1080/1369118X.2020.1748090.

Warner R and Sloan R (2013) Beyond notice and choice: Privacy, norms, and consent. *Journal of High Technology Law*. Available at: www.scholarship.kentlaw.iit.edu/fac_schol/568 (accessed 5 June 2020).

Zack ES, Kennedy J and Long JS (2019) Can nonprobability samples be used for social science research? A cautionary tale. *Survey Research Methods* 15(2): 215–227.

Zuboff S (2018) *The Age of Surveillence Capitalism*. London, UK: Profile Books Limited.