PHILOSOPHICAL TRANSACTIONS A

rsta.royalsocietypublishing.org

Opinion piece

Cite this article: Drew C. 2016 Data science ethics in government. *Phil. Trans. R. Soc. A* **374**: 20160119.
http://dx.doi.org/10.1098/rsta.2016.0119

Accepted: 26 July 2016

One contribution of 15 to a theme issue 'The ethical impact of data science'.

Subject Areas:
artificial intelligence, algorithmic information theory

Keywords:
data science, ethics, government, algorithmic accountability, public research

Author for correspondence:
Cat Drew
e-mail: cat.drew@cabinetoffice.gov.uk

Data science ethics in government

Cat Drew

Cabinet Office, London, UK

CD, 0000-0001-8485-7930

Data science can offer huge opportunities for government. With the ability to process larger and more complex datasets than ever before, it can provide better insights for policymakers and make services more tailored and efficient. As with all new technologies, there is a risk that we do not take up its opportunities and miss out on its enormous potential. We want people to feel confident to innovate with data. So, over the past 18 months, the Government Data Science Partnership has taken an open, evidence-based and user-centred approach to creating an ethical framework. It is a practical document that brings all the legal guidance together in one place, and is written in the context of new data science capabilities. As part of its development, we ran a public dialogue on data science ethics, including deliberative workshops, an experimental conjoint survey and an online engagement tool. The research supported the principles set out in the framework as well as provided useful insight into how we need to communicate about data science. It found that people had a low awareness of the term 'data science', but that showing data science examples can increase broad support for government exploring innovative uses of data. But people's support is highly context driven. People consider acceptability on a case-by-case basis, first thinking about the overall policy goals and likely intended outcome, and then weighing up privacy and unintended consequences. The ethical framework is a crucial start, but it does not solve all the challenges it highlights, particularly as technology is creating new challenges and opportunities every day. Continued research is needed into data minimization and anonymization, robust data models, algorithmic accountability, and transparency and data security. It also has revealed the need to set out a renewed deal between the

THE ROYAL SOCIETY PUBLISHING

citizen and state on data, to maintain and solidify trust in how we use people's data for social good.

This article is part of the themed issue 'The ethical impact of data science'.

1. Opportunities for data science in government

Data science can offer huge opportunities for government policymaking and service provision. With the ability to process larger and more complex datasets than has previously been possible, it can provide better insights for policymakers and make services more tailored and efficient.

The Government Data Science Partnership has been set up to harness the power of these new data, and data science techniques, and to support departments in making the best use of them. Departments have increasingly been using data science in a number of ways:

— Allowing data-led decisions by non-technical analysts/specialists, e.g. COBR (the government's emergency planning committee) now has a dynamic, interactive visualization tool that allows non-specialists to help respond to emergencies.
— Improving insight into citizens' views, needs and experiences by analysing unstructured data, such as letters, phone calls or social media, e.g. the Office for National Statistics (ONS) is exploring how to use Twitter data to understand the movements of particular populations (so people can plan local services) and the Ministry of Justice is analysing social media comments to see what people think about, and how they use, the courts system.
— Anticipating change and responding more quickly, e.g. the Government Digital Service (GDS) has created predictive models of traffic to gov.uk pages to help spot issues with pages or services more quickly and the Food Standards Agency has used Twitter data to predict norovirus outbreaks.
— Targeting and tailoring services, e.g. the Cabinet Office's Policy Lab and the joint Health and Work Unit performed a cluster analysis that revealed that people on health-related benefits had important non-health-related needs, and that these varied between different groups and required different policy interventions.
— Maximizing the use of the increasing amount of digital data that is available, e.g. the ONS is exploring how it can scrape price information from supermarket websites to feed through to economic outputs.

The Partnership has been using these 'demonstration projects' not only to show what is possible and inspire others but also to develop the data skills and capability of civil servants, and address some of the technical and legal barriers that affect data science work. The accelerator programme has, so far, given mentoring and technical support to 30 analysts to start data science projects, with more cohorts planned for this year; we are consulting on new legislation to make better use of data; and we are creating open registers of data, which can be accessed through application programming interfaces, reducing the need for bulk data transfers.

2. The need for an ethical framework

As with all new technologies, there is a risk that we do not take up the opportunities of data science and miss out on the enormous potential it can bring. Our research[1] shows that people think that it would be irresponsible not to use data science in certain cases, and that we should

[1]Primary research in this report was created by Ipsos MORI as part of the Public Dialogue for Data Science Ethics, jointly funded by Sciencewise and the Government Data Science Partnership. See https://www.ipsos-mori.com/researchspecialisms/socialresearch/specareas/centralgovernment/datascienceethics.aspx.

not lag behind other countries. Therefore, an important strand of the Partnership's work has been to create a Data Science Ethical Framework.

> I just worry what we will do if we don't do this. Other countries will – China! In that sense we don't really have a choice, we must.
>
> High Tech, event 2[2]

> And we want people in government to feel confident using new techniques. This means setting out clear guidance that brings together the relevant laws and best practice, gives data scientists and their teams robust principles to work with. It is all about encouraging new and innovative ways to better solve problems and deliver. So, today we are launching our new *Data Science Ethical Framework*, setting out in one place our framework for using data.
>
> Matt Hancock MP, Minister for the Cabinet Office[3]

There are already clear laws (for example, the Data Protection Act 1998 and intellectual property law) and professional practice that govern how civil servants work with data. But they can be complex, held in different places and not written specifically with data science in mind. While these laws are flexible, data science allows us to push the boundaries of what we have previously been able to do with data and analytical tools, and there are a few grey areas. We want people to be able to navigate the rules of how they can innovate with data confidently, and in the knowledge that they are acting lawfully.

Ethics goes beyond the law. The law codifies some elements of ethics (for example, wearing a seatbelt), but other judgements about what is right and wrong are governed by a wider moral understanding (for example, smacking a child). These ethics and understandings shift over time. Therefore, not only will technological advances create new possibilities and challenges that we have not had to consider before, but also these advances will change people's attitudes towards data.

Current behaviours show that people are more relaxed about using and sharing data under certain circumstances. For example, there are 500 million posts on Twitter every day, which are in the public domain for everyone to see. But there are other well-reported instances of concern over data. And there is plenty of research documenting concerns that people have with how their data are used.[4]

Of course, we know that people's views do not always match their behaviours. For example, nearly half of people across 20 countries say that they are willing to pay for increased levels of data privacy, but only a quarter of people, questioned in the same survey, say that they have taken free, basic steps to increase the privacy settings on their browser. This means that three-quarters of those who say that they would pay for additional privacy have not changed a simple setting on their computer.[5]

[2]London workshop that was part of the Ipsos MORI public dialogue workshops.

[3]See https://www.gov.uk/government/speeches/data-science-ethical-framework-launch-matt-hancock-speech.

[4]See Ipsos MORI (2014) Attitudes towards data sharing (for the RSS) (http://www.slideshare.net/IpsosMORI/public-attitudes-to-the-use-and-sharing-of-their-data); Demos (2012) 'The data dialogue'; Digital Catapult (2015) Trust in Personal Data: A Review (https://www.digitalcatapultcentre.org.uk/pdtreview/); Ipsos MORI and Demos (2015) Wisdom of the crowd: a guide to embedding ethics in social media research (https://www.ipsos-mori.com/Assets/Docs/Publications/im-demos-social-ethics-in-social-media-research-summary.pdf); Sciencewise (2012) Public dialogue on data openness, data re-use and data management (http://www.sciencewise-erc.org.uk/cms/public-dialogue-on-data-openness-data-re-use-and-data-management/); YouGov (2013) 'Open data' survey commissioned to support the Shakespeare Review into Public Sector Information (http://cdn.yougov.com/cumulus_uploads/document/vw2lvf25eh/Analysis-of-survey-on-%E2%80%98Open-Data%E2%80%99-commissioned-to-support.pdf).

[5]See http://www.ipsosglobaltrends.com/files/gts_2014_web.pdf.

We want civil servants not only to be confident that they are working within the law and can innovate, but also to think through how projects might be received by the public. In addition to preventing potential negative effects for citizens, projects that inadvertently raise public concern could jeopardize the future use of these tools, and we want to create some standards that mitigate this risk.

(a) Our approach

We took an open and user-centred approach to creating an ethical framework: we worked with departments to understand how they created data science projects; we convened expert roundtables to understand current and future ethical challenges with data science; and we worked with Sciencewise, Ipsos MORI and Codelegs to conduct a public dialogue to understand views on data science within government and increase awareness of the ethical challenges it raises. At each stage, we worked with data scientists internally and externally to make sure that we had understood how they approach data science projects and to ensure that we were creating something that made sense and was useful as a practical guide for everyday use.

The public dialogue consisted of three methods of public engagement:

1. A series of **deliberative public workshops**[6] to develop qualitative insight into public opinion on the appropriate use of data science. This included the development of a series of in-depth case studies and hypothetical abstracts to introduce the concept of data science and explore ethical issues.
2. Recruitment for and analysis of an **online survey**[7] to develop robust quantitative evidence on what the public thinks makes government data science projects appropriate. This included a conjoint exercise to identify the underlying factors, driving attitudes and decisions about government use of data science.
3. Use of the results of the online survey to develop a **visual interactive tool** to engage a wider audience in a public debate around data science.

Building on the discussions with external experts, by testing with civil servants and dialogue with the public, we have recently published the first draft of the ethical framework for consultation. It contains six main principles:

1. Start with clear user need and public benefit.
2. Use data and tools that have the minimum intrusion necessary.
3. Create robust data science models.
4. Be alert to public perceptions.
5. Be as open and accountable as possible.
6. Keep data secure.

It also contains a checklist of questions under each principle, with a sliding scale from very acceptable (green) to more challenging (blue). There are some departments whose public protection priorities mean that they will find themselves on the blue side (e.g. working with very sensitive information or not being able to be fully open about how they are working). The point of the framework is not to stop these types of projects (indeed, our research with the public found support for them), but that departments should consider the project carefully and see whether there are any amendments they could make that would bring them further towards the green side.

[6]Reconvened workshops were conducted with 88 people across London, Taunton, Sheffield and Wolverhampton.

[7]An online survey of 2003 people aged 18–75 in Great Britain was conducted between 24 February and 7 March 2016 using the Ipsos MORI Access Panel.

3. Each principle in turn (including supporting findings from the public research)

(a) Start with clear user need and public benefit

Data science projects should always start with a clear policy or operational need, which is an important first point for two reasons. Firstly, participants were more willing to support data science projects when there is a clear public benefit (and when they can see the value of data science over more traditional methods). And secondly, the value of the public benefit affects how much risk participants were willing to take in the design of the data science, for example in the type of data used (principle 2) or the type or volume of intended or unintended consequences that might occur within the model (principle 3). For example, participants were content for government to use sensitive data for matters of high public benefit, e.g. preventing terrorism.

Understanding how the public feels about the benefit of each project (principle 4), therefore, helps one to design a project with these risks in mind. It is also important to factor in the likelihood of the project achieving the public benefit promised, and communicating that it has been achieved.

> Depends on how big the risk is—the bigger the risk, the wider you have to cast your net. If it would be disastrous to miss them, then you have to include innocent people. Better that a few innocent people are a bit cross at being stopped, than a terrorist incident—because lives are at risk.
>
> Taunton participant.

Our research shows that the public is more likely to accept projects where there is a direct personal benefit (e.g. giving personalized employment advice), benefit to a local community, or public protection. Therefore, it is important to be as specific as possible in setting out the use case: setting out who will be benefit and how. For research or policy analysis projects, this means translating how this better understanding will benefit individuals or communities. In all cases, the public feels strongly that data should not be collected and then sold for profit or commercial gains.

People's views on public benefit are, of course, dependent on their overall attitudes to the particular policy (and government in general) and likely course of action taken as a result. Therefore, engagement on data science needs to be done alongside/through the lens of tackling the overall policy challenge, rather than the data method alone.

(b) Use data and tools that have the minimum intrusion necessary

The government should always use the minimum data necessary to achieve the public benefit. This 'minimization principle' is set out for uses of personal data within the Data Protection Act.

There are steps that we can take to do this, e.g. de-identifying or aggregating data to higher levels, keeping data in registers and querying against them (rather than seeing the whole dataset), or using synthetic data. We need to be alert to the fact that increasing the amount of data available increases the possibility of de-identified data being re-identifiable.

Often government will need to use identifiable data (and sometimes sensitive identifiable data), or include large datasets to create good algorithmic models or identify wrongdoers. In our research, participants were concerned about how this seemed to be in conflict with the principle of innocent until proven guilty and using data from people who have done nothing wrong to identify those who have. However, they did understand the need for this if it increased the effectiveness of projects with clear public benefits. Participants with more experience of government services were more accepting of using identifiable data as they could clearly see how the service could be improved by doing so.

It is important to think about how to have the minimum intrusion necessary at different stages of the project as these could vary, e.g. data collection, data interrogation and the intervention

made as a result. For example, we might need to collect everyone's de-identified data to build the model, but then only run the data interrogation with the re-identified data of those identified as risky. The public is more concerned about interventions that target individuals rather than groups, and where there are negative unintended consequences.

Social media data need to be treated with careful consideration. Legally, social media data are personal, and need to be processed fairly (within the terms and conditions of the provider). Participants had mixed views. Some do not expect government to use the data, and do not think people are aware that the data would be used in this way; others feel that it is acceptable for government to use publically available data. There is a difference in whether you use individual tweets, aggregated tweets or the metadata about tweets. It might be more appropriate to use social media data to spot trends or clusters of activity and alert local service providers to take action, than to take action yourself, but this would depend on the level of public benefit, level of consent and context in which it was provided (e.g. expectations of how a tweet is used are probably very different from sensitive discussion on Mumsnet—although both are publicly available).

In a similar way, participants could see the value of government using consumer or transactional data from outside sources where there is a clear need, but they might not necessarily expect it. This 'context collapse' causes concern and paranoia about where this might lead.

(c) Create robust data science models

Good machine learning models can analyse far larger amounts of data far more quickly and accurately than traditional methods. But these tools are dependent on the input data, and do have limits. Existing law (the Data Protection Act) states that personal data should not be used in a way that *unjustifiably* adversely affects someone. One of the key elements of how participants assessed risk for individual projects was the impact on individuals if the data science is wrong. Participants were more worried about projects where there would be an incorrect significant negative impact on a small number of people (as opposed to a small impact on a large number of people), or where there is an automated intervention following an incorrect decision (as opposed to a human investigation). This was specifically related to an example about identifying benefit fraud, and, while the responses might vary from policy to policy, it shows that balanced assessments of risk and scale and level of impact are made.

Our research shows that, in order to support a specific use of data science, the public needs to approve of both the public benefit and the value of the data science method over or alongside more traditional methods; and so it is also important to communicate this.

There are a number of things that can impact the robustness of the project.

Data. The algorithm is only as good as the data inputted to it. Using historically biased data can perpetuate or reinforce biased decisions. New digital datasets can provide real-time insight, but might not be representative (meaning that data science does not completely obviate the need to collect new data). Algorithms can search through huge datasets and make decisions based on variables, which might be acting as proxies for protected factors, such as race or ethnicity.

Models. There can also be risks in creating data science models. Data science can gain new insight from existing data, rather than collecting new data to answer each new question. Some of these data will be inferred, and, where these are sensitive (for example, health conditions from shopping habits), extra care is needed.

Machines make trade-offs to find their target answer; either by narrowing down criteria or groups, but not identifying all data points that meet that criteria (false negatives), or by widening out criteria or groups, but identifying some incorrect data points (false positives). The public is concerned not only about missing false negatives when this affects public safety, but also about including false positives when there is an unfair negative impact for that person.

Finally, it is difficult to translate a complicated policy into a coded algorithm to make operational decisions (for example, benefit applications).

It is really important to be honest about the level of confidence in the insight as this will affect the decision made on the back of it, and to keep track of the provenance of the data. Policymakers

or operational staff should work together with the data scientist to iterate the model, check that it is working well and be able to explain any unintuitive anomalies.

In summary, different machine learning techniques work best with different projects. Automated machine learning works best when the policy can be easily translated into an algorithm, when there is high confidence in the data and when there is clear recourse for a decision made about an individual. In other cases, there should be more human oversight, and the result should be used with other insight to inform the decision.

(d) Be alert to public perceptions

As discussed earlier, ethics are society's collective moral understanding of what is the right thing to do. Some of this is codified in law, and some of this is not. Ethics become more important when advances in technology are pushing our understanding of the law to its limits. Public perceptions are more diverse, and depend on a multitude of factors, such as current events and media presentations of risk.

Both the law and ethical practice require us to understand public perception so we can work out what we should do.

The framework sets out how we need to balance public benefit (principle 1) with the risks to privacy (principle 2), validity and unintended consequences (principle 3). Understanding public perceptions about the benefits of the project helps us to make these judgements and to shape the research design. As public perceptions of the data science project are so highly linked to attitudes to the overall policy, it will also help to identify whether they are related to that or the data science method *per se*.

More specifically, the law (the Data Protection Act) states that the use of personal data for operational (i.e. non-research) purposes has to be 'fair, proportionate and compatible with the original purpose for which it was collected'; to work that out we need the understanding of how people would reasonably expect their personal data to be used.

The government supports open approaches to policymaking, and many are set out in the Open Policy Making toolkit. Policy Lab is a small team based in the Cabinet Office that supports policymakers to put citizens at the heart of how they design policy, and works with them to understand their needs and opinions and to co-design solutions. Data science techniques (including social media analysis) can also play an important role in helping us to understand public perception.

(e) Be as open and accountable as possible without putting people at risk

Transparency is essential to make the case for the benefits of data science and to avoid accusation of nefarious 'secret' big data projects. It is also a good antiseptic for unethical behaviour. Ideally, people would like transparency at all stages of a data science project, being told when and why data are being collected about them as well as whether the outcome of the data science project is achieved.

> I agree it's about being open and upfront about what the outcome of gathering the data has been – that's the bit we don't hear about! We don't know what changes have been made in Taunton as a result of the data being collected. None of us know the impact our data collection has had on the local community. As a taxpayer, how is that money being spent?
>
> General public, Taunton event 1

However, there are some occasions when it is rightly not possible to be fully open, as doing so would jeopardize the project goal by, for example, allowing people to understand how we are identifying illegal activity and game the system. Participants in our research understood these instances in principle.

There are other areas where it is currently more challenging to be fully transparent, but we need to find ways to achieve this. Publishing the algorithm in programing language alone is not sufficient, so we need to find ways to explain in plain English how machines make decisions. Where algorithms are changing, depending on the data put in, we need to be able to track which version of the algorithm made the decision. What is currently harder is for 'black box' algorithms where the decisions that the machine makes to reach a goal are unknowable.

Our research found that full transparency is not always necessary as long as the workings of the computer are known to staff, but that decisions are accountable. Participants needed to know when decisions have been made about them by machines, be able to find out how that decision was made (in plain English) and be able to challenge that decision and obtain recourse if that decision was incorrect. Therefore, we also need to be open about the risks, what we are doing to mitigate them, and how we are putting things right if they go wrong.

Oversight throughout the process is critical: in terms of the way the project adheres to professional standards and frameworks and—critically for the public—how there is an element of human oversight over machine decisions. Our research found that the public is extremely wary of putting too much trust in machines. Humans have an important role to play in the interpretation of data and turning machine-generated insight into knowledge and action.

> People should be involved in decision making when the risk is related to human beings or businesses. Don't want to become too reliant on machines or 'blame' algorithms for decision making. General public, Taunton event 2

(f) Keep data secure

We know that the public is justifiably concerned about people's data being lost or stolen. For participants involved in the public research, security was a basic and obvious point. Government has a statutory duty to protect the public's data, and as such it is vital that appropriate security measures are in place. There is much guidance (e.g. the Data Protection Act and the government's own Security Policy Framework) to set out how data should be collected, stored, shared, processed and deleted.

Government is creating infrastructure which will hold data more securely. The GDS is creating registers of data which can be queried against, rather than drawing together bulk datasets. The Administrative Data Research Network has created a number of 'safe havens' where administrative data (data routinely collected by government) can be anonymized and linked, with strict controls over who has access to the data and for how long.

4. Discussion

The open policymaking approach through which we have developed the framework—involving both technical experts and members of the public—has revealed some important directions for future communication and data policy development.

(a) Engaging with data science

The public research both tested the framework concepts and revealed the process through which people make their considerations. This gives us useful directions for how we engage in the future.

The research showed that data science examples can increase broad support for government exploring innovative uses of data. But that only goes so far. Participants' attitudes are highly context driven. On a case-by-case basis, they consider acceptability in two stages. Firstly, they consider the overall policy aim and likely intervention as well as whether data science provides more value than more traditional methods. Only if they accept that can they then move on to a nuanced risk assessment of balancing the level of public outcome and project efficacy against privacy and unintended consequences.

Participants agreed with the principles set out in the framework, spontaneously using similar language. But there are no absolute red lines; as described above, decisions are taken very much in the context of the overall policy area.

There was, perhaps unsurprisingly, very little awareness about data science and how much people generate data in general. But participants were unconsciously experiencing data science applications every day. This is important as, if people were not aware of the potential benefits of data science, they could not understand or were sceptical about how it worked, and therefore they could not perform a risk assessment. Therefore, we need to increase the public's data literacy as well as that of the civil service, and find new ways to engage with people who are unaware of how or where their data are being used.

Revealing case studies was, therefore, critical to getting the public to engage. As Rempel points out,[8] the choice of these was important in framing how people react to data science. We chose a range of case studies—but as we know how much attitudes to the overall policy area have impact on the reaction to the data science project, people were also giving their views on general policy topics. That this is hard to disentangle is important. Data scientists need to work with policymakers and operational members of staff to engage the public in (and understand their attitudes to) the overall policy area, as the level of value will help to determine the data and method used.

A consultation workshop[9] unearthed an interesting point about the purpose of engagement on data science projects. Engagement takes time and resources, and could potentially decrease trust if too many hypothetical risks are highlighted. For example, in the public research, high technological literacy did not necessarily translate into support for data science methods, but instead sometimes suspicion about their use. Rather than engaging to build trust in the data science method *per se*, engagement might be better used to understand how people value the overall policy area (and whether they would be happy for their data to be used to solve it) and be inherently trustworthy.

Inherent trustworthiness is important as not everyone will want to engage, but everyone will want to know if government is using their data appropriately.[10] Publishing the Data Science Ethical Framework is an important first step in this. But it alone does not solve all the challenges that it highlights, and it has revealed the need for work around improving trust in data more generally, not just in government but also in the private, academic and civil society sectors. This is set out in the section on data policy development below.

To summarize around communication, the dialogue has suggested four main points for us to consider:

1. Highlight positive case studies to show the benefits of data science.
2. Take open policymaking approaches to engage people in the problems themselves (not just the data science method).
3. Increase data literacy, so people can understand the benefits of data science above other methods, and so they can assess risk.
4. Communicate what we are doing to mitigate risks and be trustworthy.

The conjoint survey allowed us to create four segments of the population with varying attitudes to data science: data adopters, adapters, pragmatists and wary. Communication techniques will vary from segment to segment, and we should be thinking about how we make the best use of advocates within the public who can make the case for data science. Proximity to a tangible public benefit is also highlighted here. The group of people with a high level of data

[8]Rempel E. 2016 The problem of public engagement, public policy & public data. Institute for Policy Research, 19 May. See http://blogs.bath.ac.uk/iprblog/2016/05/19/emily-rempel-the-problem-of-public-engagement-public-policy-and-public-data/.

[9]This was a consultation event hosted by Nesta on 19 May 2016 to launch the framework in which stakeholders explored the public research and how to update the framework based on this.

[10]Stilgoe J, Irwin A, Jones K. 2006 The received wisdom: Opening up expert advice. London, UK: Demos.

interactions with government were the most supportive of data science, as they could see how the services that they regularly experienced could be improved as a result.

(b) Further technological and data policy work

But there is further work to do.

The ethical framework is a start, but it does not solve all the challenges it highlights. Continued research is needed into technological infrastructure and methods to ensure data minimization and anonymization, to create robust data models, allow algorithmic accountability and transparency and to keep data secure. This is not something that the UK government is tackling alone; the US government has just published *Big Data: A Report on Algorithmic Systems, Opportunity and Civil Rights*[11] to highlight its work to tackle the challenges around discrimination and human rights posed by machine learning.

Algorithms and machine learning techniques pose big questions around accountability and oversight. How can we achieve true transparency for algorithms written in code that few can understand? Or where constantly changing input data or their 'black box' nature means that it is difficult or even impossible to decipher how algorithms reach decisions? Or where—for security or proprietary reasons—algorithms cannot be published. There are many different forms that oversight could take. Nesta recently made the case for a Machine Intelligence Commission to ensure that the public interest is protected as a new generation of algorithms is developed. The Alan Turing Institute has recently been set up as the UK's national institute for data science. The government needs to bring together various experts in this field as it considers the Science and Technology Committee's recommendations for a Data Science Ethics Council. In the USA, the Council for Big Data Ethics is starting to create some *thought leadership pieces*, including on how to address disjuncture in research ethics across the fields that data science brings together, and industry where iterative algorithmic insight generation is considered differently from academic research ethics. As with the ethical framework, it will be important that forthcoming work is theoretically grounded, but written in a practical way so that data scientists, policymakers and others can implement innovative uses of data now, confident that they are working appropriately.

Many of the issues raised in the public dialogue concern the use of data more generally, as well as data science. We are at an exciting moment for data-driven government. There is an opportunity for us to set out a new relationship between the citizen and the state which shows what data we use, why (the reciprocal benefit) and how we are managing risks.

New technology means that we can offer citizens greater visibility and control for where their data are used for transactional services. Although at the very beginning of the journey, we are starting to explore how citizens can see and control how their data could be shared across the government (for example, viewing driving licence information and giving drivers opportunities to port the data to outside organizations, e.g. car hire companies) and private sectors—initiatives are starting to more clearly build consent into services and even share the value of the data back with customers.

But government is not just about providing transactional services. There are also personal data uses which have wider public benefits: to identify and intervene with people, to carry out aggregate-level research and policy analysis to make better decisions and plan services, and to provide anonymous or non-human open data which the public can use to innovate or hold government to account. For these areas, consent is not the appropriate mechanism; holes in data mean that we might not be able to identify vulnerable people or wrongdoers or that we may not have complete datasets for policy research that helps us plan services, nor for the open data that businesses can use to innovate and the public can use to hold us to account. For these areas, it is crucial to continue the efforts described above to ensure trustworthy practice, security and oversight.

[11]See https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.

Finally, we cannot stand still. Technology is changing, and we need to keep pace with it. Artificial intelligence provides huge opportunities for deep learning, and blockchains offer the promise of greater accountability and security. The biggest risk would be not making the best use of the data and tools that are increasingly becoming available. This means scanning the horizon for new opportunities as well as spotting potential challenges, and quickly convening networks of experts to overcome them. The Data Science Ethical Framework has been published as a first iteration. It needs to be a document that is robust in providing a coherent standard across government, flexible so it can adapt to new data and technology as they appear, and written with data scientists in a practical way so that people can confidently find innovative ways to put data to the best use.

Competing interests. The author declares that there are no competing interests.
Funding. I received no funding for this study.