Data policy and the public: shaping a deeper conversation

Josephine Suherman-Bailey

January 2015

You can comment on this report here.

This report, and the others in the series, has been produced by Sciencewise.

Sciencewise is a BIS funded programme to encourage the more widespread use of public dialogue in policy involving science and technology. Sciencewise provides advice and guidance to help those involved in the development of policy to understand and to take into account the views and values of the public in the development of policy involving science and technology. Sciencewise is able to provide:

- Advice and guidance on public dialogue and engagement.
- Assistance with the implementation of engagement as appropriate
- Financial support for the implementation of selected public dialogue projects
- Training and mentoring to assist those involved in policy development to build their understanding of the benefits and their confidence around engagement with the public.

If you would like to find out more about Sciencewise and the support we can offer, you can:

- visit our website: http://www.sciencewise-erc.org.uk/
- follow us on twitter: @Sciencewise
- contact us by email enquiries@sciencewise-erc.org.uk or phone 01235 753 645

Table of Content

1 Executive Summary

The policy issues around data are complex and controversial, and are becoming more so all the time as the technology rapidly advances. There is increasingly a consensus that the public voice needs to be heard alongside others as part of the policy process. But what exactly does this mean? Where could the public voice fit in?

This thought leadership piece is designed for policy makers involved in data issues, who are considering how to bring in the public voice. The paper aims to support policy makers on how to structure their thinking on the issues, scopes a range of possible questions that could be explored and clarifies the value that public dialogue on data may bring.

Policy makers are currently facing a number of key, cross-cutting issues in policy related to data:

- For what purposes should government be able to collect, process and share data? Only the public good and improvement of all citizens' lives and services? Or also in ways that could lead to sanctions for some individuals?
- What level of choice and consent should the public have over the collection, sharing and processing of their data by government?
- What privacy concerns does the public have regarding the greater use of data? How far and in what ways are the public willing to "trade off" privacy for the benefits that greater use of data and data science can bring (known as reciprocity)?
- What conditions, safeguards and penalties for misuse does the public want to see for the government using their data?
- Should policy makers make decisions based on predictive techniques? What are the advantages and disadvantages?
- What level of transparency does the public expect from government on their use of data? What are the dangers of transparency? What are the dangers of a lack of transparency?

This paper looks at where and in what ways the public voice could add value in policy making, and what kind of problems the public might be able to help policy makers solve.

The paper is not meant as a comprehensive plan for engaging with citizens on the issues raised by the greater use of data, nor does it cover all the issues that greater use of data throws up. The paper's intention is to introduce the ways in which citizen voice can help policy makers to explore the issues, and understand how best to move forward.

Policy makers can make use of Sciencewise as a resource to help think through the options on the public dialogue tools available to them, and find more information on public dialogue on the Sciencewise website.

2 Introduction

Greater use of data by government can offer huge opportunities to create insights, which can lead to better policymaking and better public services. However, as the technology for collecting, using and analysing data moves forward at lightning speed, people are starting to think about the ethical considerations which go beyond what the law and the technology permits us to do.

There are a number of ethical issues facing government stemming from the greater use of data. These centre around trust, transparency, security, accountability, privacy, ownership, control and consent. The law - namely the Data Protection Act 1998 and the Human Rights Act 1998 - is subjective. This permits policy makers to carry out actions which are legal but whether those actions are ethical is less clear cut. And the law is not enough when the technology allows the collection, manipulation, analysis, layering and creation of insights from data in a way that wasn't possible before.

Although the public's attitudes to data issues have not been fully explored yet in the literature, what we do know is that the public has concerns around data privacy, a lack of trust in government's ability to keep our data secure, and concerns about losing ownership over their data. Policy makers could consider building on existing research by engaging the public in a dialogue when they are addressing specific policy issues of interest to the public.

Since data is an area in development and the technology is advancing rapidly, many of the terms are contested. This piece discusses data in generalist terms and refers to open data, big data and data science, areas which overlap and are defined in different ways in different quarters. Below are some definitions to work with:

Open data

"Open data is data that is made available by organisations, businesses and individuals for anyone to access, use and share."[1] (Open Data Institute)

Big data

"Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is big data."[2] (IBM)

Data science

"Data is increasingly cheap and ubiquitous. We are now digitizing analog content that was created over centuries and collecting myriad new types of data from web logs, mobile devices, sensors, instruments, and transactions...

At the same time, new technologies are emerging to organize and make sense of this avalanche of data. We can now identify patterns and regularities in data of all sorts that allow us to advance scholarship, improve the human condition, and create commercial and social value...

[1] Open Data Institute (2014) 'What makes data open?' http://theodi.org/guides/what-open-data
[2] IBM (2014) 'Big Data at the Speed of Business' http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

Virtually every sector of the economy now has access to more data than would have been imaginable even a decade ago. Businesses today are accumulating new data at a rate that exceeds their capacity to extract value from it. The question facing every organization that wants to attract a community is how to use data effectively — not just their own data, but all of the data that's available and relevant.

Our ability to derive social and economic value from the newly available data is limited by the lack of expertise. Working with this data requires distinctive new skills and tools. The corpuses are often too voluminous to fit on a single computer, to manipulate with traditional databases or statistical tools, or to represent using standard graphics software. The data is also more heterogeneous than the highly curated data of the past. Digitized text, audio, and visual content, like sensor and weblog data, is typically messy, incomplete, and unstructured; it is often of uncertain provenance and quality; and frequently must be combined with other data to be useful. Working with user-generated data sets also raises challenging issues of privacy, security, and ethics.

The field of data science is emerging at the intersection of the fields of social science and statistics, information and computer science, and design."[3] (The UC Berkeley School of Information)

3 Scene-setting

a. Why is greater use of data a good thing for government?

Greater use of data and data science opens up new ways to improve government policy and public services. It can help to:

- Enable data-led decisions by non-analysts
  - *For example, creating dynamic, interactive visualisations which can be explored by non-analysts to spot trends and patterns.*
- Understand citizen views and experience
  - *For example, analysing unstructured data such as letters, phone calls or social media to improve insight about citizen's needs.*
- Anticipate change and respond more quickly
  - *For example, real-time tracking and predictive modelling of traffic to gov.uk pages could help spot issues with pages or services much more quickly.*
- Target and tailor services
  - *For example, segmenting service's users means government can give citizens the services they need and reduce waste.*

b. What do we already know about the public's attitudes to data?

Sciencewise's Social Intelligence pieces bring together the existing research on public attitudes to open data and big data. Below are some of the key points from the big data piece, which speaks to some of the views and concerns citizens have across the data debate:

[3] The UC Berkeley School of Information, 'What is data science?' http://datascience.berkeley.edu/about/what-is-data-science/

- When asked, the public are ostensibly opposed to any form of data use and collection by government and companies, but in practice the public consider there to be no alternative to sharing personal information with government and companies in the modern world and expect it to increase in future[4].

- Personal benefit is the strongest incentive for being in favour of the collection and use of personal data by government and companies, but the public report currently seeing little benefit from sharing their data and little confidence that they will see benefits in future. The public also identify public goods (e.g. health research, prevention and detection of crime, and unearthing of dishonesty or fraudulent behaviour) as potential benefits of personal data use[5].

- The public is particularly concerned about losing control of their personal data, with fear that they will become a victim of fraud or identity theft, and that their data will be shared with others without their knowledge or agreement[6].

- Offering a specific personal or public benefit can significantly increase the general public's acceptance of the collection, sharing and use of their data by government and companies, but even when a specific benefit is offered, the public remain concerned about the collection, sharing and use of particular types of personal data (e.g. bank account, savings and pension details) [7].

- There is no consistent "public view" on what constitutes personal data, the benefits of sharing personal information and behavioural data, and comfort levels with different uses of data. The public can be segmented into a number of groups sitting along a continuum between pro- and anti-sharing[8].

- The public thinks that personal data should only be used by government and companies for their personal benefit. People are keen to have more control over the use of their personal data and want stronger safeguards towards its use, and there is strong support from the public for more information on how government and companies collect, share and use data[9].

Whilst the Sciencewise pieces on open data and big data give some insight into public opinion, there has not been nearly enough research into the topic. There has been very little done on attitudes to different specific uses of data science, which throw up very complex and controversial issues. Due to the nature of the issues, a more deliberative and participatory approach should be considered, to bring in the public voice on data.

c. Why should policy makers bring in the public voice on data?

The technology for collecting, using and analysing data is developing rapidly, and the potential and possibilities for the use of data has grown. Whilst this is an exciting time for policy makers,

4 Sciencewise, Social Intelligence: Big Data (2014) http://www.sciencewise-erc.org.uk/cms/assets/Uploads/SocialIntelligenceBigData.pdf
5 ibid
6 ibid
7 ibid
8 ibid
9 ibid

the developments throw up ethical questions which need to be carefully considered, and policy makers should consider bringing in the public voice to help them:

> "Mechanisms for hearing the views of citizens and other stakeholders should be institutionalised for all policy decisions. However, some policy decisions warrant and require a broader or deeper level of engagement with citizens than others. This is particularly true… on issues that will likely impact or concern citizens in a significant way. For such issues, it can be advantageous to use deliberative methods of citizen engagement, whereby groups of citizens engage deeply with the issue in question." (Open Government Guide)

Data is certainly an area which will, in the words of the Open Government Guide, "impact or concern citizens in a significant way"[10], and there are a number of reasons why policy makers should consider bringing the public voice in:

1. **Greater transparency and legitimacy.** Public trust is crucial to the success and ability of government to continue to make the most of the opportunities presented by data science. Policy makers demonstrating that they have engaged the public in a debate on complex and controversial issues can increase the legitimacy of policy. Publishing responses to input increases transparency, and builds public confidence that government decisions are based on appropriate criteria and evidence[11].

2. **Bringing in the public voice can increase trust and avoid the public rejection of policy.** Implementing difficult decisions depends on citizens' consent and support. The incorrect use of data could have negative consequences for the government. Unless citizens understand and are engaged in the decision themselves, trust is easily lost (OECD, 2009)[12]. Involving the public can give people a sense of ownership over the final decision, lowering the likelihood that it will be challenged or rejected.[13] Particularly given current levels of public mistrust of government on data issues, it is important for policy makers to be attuned to the public's views on this issue.

3. **Citizens can help policy makers to understand the ethical issues around the greater use of data.** Acting in an ethical way goes wider than the law and may require additional oversight. There are strong data protection and privacy laws in place, but many of the terms employed by the law are subjective, and individual members of the public have different moral stances which vary according to context and shift over time. Public dialogue can help policy makers understand the public's priorities and ethical concerns.

4. **It is difficult for policy makers to act with confidence on complex and controversial issues when they are unclear on the public's outlook on the issue.** The greater use of data and data science could create opportunities for innovation which can lead to more tailored, responsive and better government. However, the lack of a clear steer or sense of

[10] Open Government Guide: Engage citizens in deliberation on a priority issue
http://www.opengovguide.com/commitments/engage-citizens-in-deliberation-on-a-priority-issue/

[11] Open Government Guide: Engage citizens in deliberation on a priority issue
http://www.opengovguide.com/commitments/engage-citizens-in-deliberation-on-a-priority-issue/

[12] OECD, 2009, Focus on Citizens: Public Engagement for Better Policy and Services

[13] Open Government Guide: Engage citizens in deliberation on a priority issue
http://www.opengovguide.com/commitments/engage-citizens-in-deliberation-on-a-priority-issue/

legitimacy stemming from knowledge of the public's outlook on an issue can cause policy makers to be overly cautious or risk-averse, which risks the gains we could make from greater innovation.

5. **Bringing in the public voice can help policy makers to identify questions or issues they hadn't thought of.** Some of the questions and issues it would be useful to engage citizens on are outlined below; however, sometimes the most useful part of engaging citizens is discovering what you didn't know. Eliciting the views, values, knowledge and experiences of the public can offer new perspectives on issues and be the source of important new information and ideas[14], leading to better decision making.

4 On what issues could policy makers benefit from the public voice?

This section explores some of the issues raised by greater use of data, how policy makers could benefit from bringing in the public voice on these issues, and suggests some key questions policy makers might want to put to the public.

a. The purpose of data sharing

Key questions for public dialogue:

- For what purposes should government be able to collect, process and share data? Only for the public good and improvement of citizens' lives and services? Or also in ways that could lead to sanctions for individuals, although this might have a wider public benefit? For example:
  - To improve public services by tailoring to the individual so they receive the right offer at the right time?
  - To improve public services by improving efficiency?
  - To help people in debt receive tailored advice?
  - To target public service messages?
  - For social and economic research and statistics purposes?
  - For medical research and statistics purposes?
  - To identify and prevent fraud?
  - To identify and prevent anti-social behaviour?
  - To identify and prevent illegal immigration?
  - To identify and prevent terrorism?
  - To restrict welfare benefits or services to particular individuals or groups?

One of the core issues government is concerned with is deciding for what purposes it can use data. Government could use data for the public good and improvement of citizens' lives and services only. For example, to improve public services by tailoring to the individual so they receive the right offer at the right time, to improve public services by improving efficiency?
To help people in debt receive tailored advice, for social and economic research and statistics purposes, to identify and prevent fraud, to identify and prevent terrorism or to restrict welfare benefits or services to particular individuals or groups.

[14] Ibid

Some civil society campaigners feel very strongly that it is unethical for improvements in technology for collecting and analysing data to lead to sanctions for individual citizens, even if an argument around the wider public benefit can be made. This is one area in which the public voice could help to clarify the issue.

We do have some understanding of the public's outlook on these issues. The Wellcome Trust (2013) found the main benefits identified by members of the public for the collection and use of personal data by government to be: 'the Government identifying needs, planning resources and services, and allocating funds'; 'prevention and detection of crime and, including terrorism'; 'identifying social/population trends and statistics'; 'unearthing dishonesty (e.g. fraudulent benefit claimants and tradesmen)'; and 'availability of vital medical information in a medical emergency.'

In the public sector, some participants in the Wellcome Trust's (2012) focus groups stressed the importance of the collection of health data when it benefits the individual. Data sharing within the NHS was considered by those in these focus groups to be positive, with the perception that more data sharing could be done within the NHS (Wellcome Trust, 2012). When offered a specific public good, acceptance of data sharing can increase significantly. Ipsos MORI, for example, found that:

"People are most supportive of individuals' data being used when there are tangible public service benefits. Nine-in-ten (88%) support the use of people's data to help develop treatment for cancer, three-quarters (73%) support data being used to improve the scheduling of transport services and seven-in-ten (70%) support data use to prevent crimes"[15]

As well as to accrue benefits to individuals and the public, the use of personal data for enforcement is also considered by some participants in opinion polls and public dialogues to be a benefit. For example, the detection of fraud is mentioned by participants in a number of studies as being a benefit of data collection and use by government and companies, for example in identifying benefit cheats. The Wellcome Trust (2013) report that some think more could be done to catch those who flout the system by linking data between organisations (e.g. linking Facebook data and benefits payments).

However, others are concerned about data being used by government to punish or withdraw a benefit or service from individuals (Wellcome Trust, 2013). When talked through a range of scenarios of how data could be linked between organisations (public and private) for a potential personal or public benefit, participants in the Wellcome Trust's (2013) focus groups were typically concerned about data being used to target specific individuals or groups of individuals.

The government is already looking at some of these issues through its data-sharing work. They are keen to explore whether some of the barriers to sharing and linking different datasets in government can be removed in order to develop a better understanding of the economy and society, deliver more targeted and joined-up public services, and save public money lost through fraud, error and debt. Government departments are engaging with civil society through an open policy making process to explore the benefits, risks, limitations and governance for sharing personal data within government.[16]

The public's default position when it comes to the sharing and linking of data within government and public services is one of significant caution, for many of the reasons outlined in the following section. However, while the public is not willing to give government a free pass to collect, share and use personal data, it is (as will be discussed in section 4.2) willing to give

[15] Ipsos MORI (2014) Public attitudes to science 2014. 'Attitudes to Big Data' section
[16] Read more at http://datasharing.org.uk/

conditional consent in particular circumstances when there are clear personal and/or public benefits on offer. Public dialogue might be a valuable way of exploring what those circumstances look like in greater depth.

b. Choice and consent

Key questions for public dialogue:

- What level of choice and consent should the public have over the collection, sharing and processing of their data by government?
- If citizens are allowed significant choice, should they have the option of 'dynamic' consent – the choice of different levels of consent for different data? Should they be able to opt in or opt out of the collection of their data? What are the advantages and disadvantages of an opt-in/ opt-out system or 'dynamic' consent?
    - If many people decide to opt out of allowing their data to be collected for research purposes, this skews the research sample and findings will not be reliable.
    - Are people making an informed or thoughtful choice if we make opting out the default?
- What is 'informed consent'? Is our understanding of it strict enough?
    - Is it possible to give informed consent when data agreements are lengthy and seldom read before they are agreed to?
    - Is it possible to give consent when people cannot easily engage with intangible notions of further use, even if it is stated?

Choice and consent are at the core of ethical issues around the greater use of data for government. Policy makers could benefit from a better understanding of the public's outlook on the ethical level of choice and consent over the collection, sharing and processing of their data by government.

A dialogue on choice and consent could look at 'dynamic' consent - different levels for different data uses - which many private sector organisations offer: "within those use cases there will be a series of incentives and you can go through and grant and revoke a percentage of these use cases, according to your comfort level".[17]

An public dialogue could also look at the public's view of the difficulties thrown up by 'dynamic' consent. For example, if a significant number of people decide to opt out of allowing their data to be collected for research purposes, the result will be a skewed and therefore unusable and unreliable research sample. On the other hand, whether we make opting out the default when collecting data could throw into question whether citizens are making an informed or thoughtful choice when they decide to opt in.

There are ethical questions thrown up by citizens having the choice to give differing levels of consent, in terms of services received. If the government is using data to tailor public services to the individual, but there is a choice to opt out of this greater tailoring of services, there could be an ethical issue around citizens receiving different levels and quality of care.

[17] Hull University Business School (2014)

Another question is around government's understanding of 'informed consent', and whether this understanding is strict enough. The ease with which organisations are able to achieve 'informed consent' has given rise to questions around the extent to which consent is or can be classed as informed. There are arguments that that informed consent as it's currently stated is no longer good enough since data agreements are lengthy and seldom read before they are agreed to, and because people cannot engage with intangible notions of further use even if it is stated.

Public dialogue could play a role in giving policy makers an insight into the level of consent the public feels is acceptable for the use of data, the advantages and disadvantages of an opt-in/ opt-out system or a system of 'dynamic' consent, and the ways in which policy can ensure that citizens understand and are really in control of how their data is used.

c. Privacy

Key questions for public dialogue:

- What privacy concerns does the public have regarding the greater use of data? How is privacy defined?
- How far and in what ways are the public willing to "trade off" privacy for the benefits that the greater use of data and data science can bring (known as reciprocity)?

Concern about privacy is the thread that runs through much of the public dialogue work so far on data, and is of vital importance for policy makers to have a handle on, as it could represent a barrier for future data science projects: Ipsos MORI (2014) found that one of the top reasons for the public opposing data use was that 'People have a right to privacy' (32%)[18]. Policy makers could have a dialogue with citizens to understand how they define privacy, and the specific privacy concerns the public has regarding the greater use of data.

The legal challenges that data science may create are around those that apply to personal data; policy makers may want to use data for purposes other than for which it was originally collected, or the combination of data sets (where data has been aggregated) could allow an individual to be re-identified. There is a legal framework, within which government needs to act. The Data Protection Act can allow departments to use data for purposes other than for which they were originally collected where this is 'fair'. Fairness will depend on the circumstances, taking into account, amongst other things, what the individual could reasonably have expected when s/he provided the information and whether the proposed use would adversely affect him/her. It is clear that government should not publish personal data that could identify or potentially identify individuals.

Individuals have a range of moral views on data and privacy, which are wider than what is codified in law. Concepts of privacy are also shifting over time, as is the ability of data technology to challenge the viability of the concept, meaning legislation – even if it were able to be more objective – would not be able to keep up with them. We are spending more of our lives 'onlife' which is transforming how we develop and present ourselves online, and how we interact with others in our information world. Although sometimes contested, evidence from

[18] Ipsos MORI (2014) Public attitudes to science 2014. 'Attitudes to Big Data' section.

some public surveys suggests the public is coming to the view that companies and governments holding lots of data about us is inevitable (if not acceptable), and people are surprised that government does not just use the data it already holds about us more.

There is also a question around how far and in what ways are the public willing to "trade off" privacy for the benefits that data science can bring. Opinion surveys suggest some willingness on the part of the public to trade-off their concerns against the potential benefits to themselves or the wider public. The evidence suggests that there is a greater willingness to give data to government departments when general benefits are offered. 68 per cent of respondents said that they would be happy to provide details to government departments if it meant that they would provide a better service, compared to 25 per cent who would not. The case is particularly stark for explaining the benefits of handing over personal details to government when looking at those who would typically be in opposition; if it means a better service, 49 per cent of this group would shift from being "rejectors" to "acceptors". When asked to make some specific trade-offs (with a clear personal or public benefit), the public appears to be much more comfortable with their data being used. When faced with specific and tangible scenarios and benefits of data being shared across public services, respondents seem much more comfortable with their data being shared. For example, 91 per cent of IIPS (Institute for Insight in the Public Services) respondents agreed with the proposition that medical staff across the country should have access to their GP medical records, meaning that their medical history would be available to services if they needed medical care outside of their area. [19]

Public dialogue on privacy could enable policy makers to achieve more nimble and public oversight to understand and reflect the current 'privacy/public interest' debate and to take into account the range of ethical stances that sit outside what is codified by law. Public dialogue could also help policy makers understand what the public understands privacy to be. Public dialogue can also give policy makers an insight into the ways in which policy can be presented to the public, and to understand how far and in what ways the public are willing to "trade off" privacy for the benefits that data science can bring.

d. Conditions, safeguards and penalties

Key questions for public dialogue:

- What conditions and safeguards does the public want to see for the government using our data?
- What penalties do people want to see for the misuse of data?

An understanding of the conditions, safeguards and penalties for misuse the public wants to see for the government using our data will be useful to policy makers in helping them to make the case for greater use of data.

The government's independent Shakespeare Review (2013) found that a significant number of respondents held favourable views regarding the release and publication of Public Sector Information (PSI). The research surveyed normal citizens as well as those who work in open data, and respondents held these views irrespective of their existing use of, or interest in, PSI and open data. 83% of respondents would approve of a general policy for open data if certain conditions - privacy, openness and security – were recognised as fundamental components.

[19] Institute for Insight in the Public Services (2008) 'Data and Privacy: How concerned are citizens about data sharing in the public services?

However, loss of control over personal data and information is found to be a significant concern of the public across surveys, interviews and focus groups. Ipsos MORI (2014) found that one of the top reasons for the public opposing data use was concern about 'abuse of personal information/identity theft' (40%)[20]. Likewise, the top five risks identified by respondents to the Eurobarometer (2011) survey were all linked to losing control of their data, including being a victim of fraud (65%), being at risk of identity theft (56%), information being used without their knowledge (34%), information being shared with third parties without their agreement (33%) and information being used in different contexts from the ones where it was disclosed (23%).

The risk of personal data theft and misuse is consistently found to be at the top of the public's list of concerns. For example, Ipsos MORI report from their public dialogue on the use of government administrative data for research, that:

"Personal data security was very important to participants, and this framed much of the discussion. They were particularly concerned about identity theft, and personal data being sold on to other organisations"[21]

More data is being generated by us and things around us, and then collected by organisations as big, unstructured and/or open (freely available) data. The ability to completely anonymise personal data is subject to debate, and may even become impossible, and these additional datasets and tools can be used to infer identity[22].

Public dialogue could play a role in supporting policy makers to understand the conditions, safeguards the public want to see for the government using our data, and the penalties for misuse. An understanding of the strong, robust safeguards the public wants to see will give policy makers the confidence to undertake data science projects.

e. Predictive techniques

Key questions for public dialogue:

- Should policy makers make policy decisions based on predictive techniques? What are the advantages and disadvantages?
- Advantages include:
    - Using data large, unstructured, real time, and social media data it was difficult to analyse before, in order to inform policy and provide a strong evidence base
    - Getting different/ better insights from data
    - The potential to target and tailor services
- Challenges could include:
    - Political bias
    - Unintended discrimination
    - Predictive bias

Some important questions for government are around predictive techniques. The advantage of using predictive techniques is the ability to build a much more rich and detailed evidence base for policy making, yet there are a number of challenges around the possible unintended

[20] Ipsos MORI (2014) Public attitudes to science 2014. 'Attitudes to Big Data' section.
[21] Ipsos MORI (2014) Dialogue on data: Exploring the public's views on using administrative data for research purposes

[22] The Royal Society (2012) Science as an Open Enterprise

bias and discrimination that can be built into predictive techniques. This is an area in which the public understanding and awareness will probably be low, and yet their outlook on how to solve these ethical issues could be very useful to policy makers. If increasing amounts of data about ourselves and other agents in our world of information (e.g. predictive techniques) are increasingly used to develop or deliver policy, there are some ethical challenges to consider.

There is an increasing range of accessible data sources, cheaper technology to store and process that data, and powerful tools for analysing it, which throws up new challenges for acting ethically, and offering transparency and accountability.

Whilst some have argued that predictive techniques have many advantages over human decision making as they can do so at scale, accuracy and without cognitive bias, and "with enough data, the numbers speak for themselves"[23], others have challenged data fundamentalism (the notion that correlation always indicates causation)[24]. Government should be aware of the risk of building in potential (and unforeseen) bias in predictive techniques.

Predictive techniques are not free from political bias. The way the algorithm is created; its desired outcome, the categories that it uses and the criteria for relevance which it uses to sort information are selected and agreed by humans and based on a particular world view[25]. They are essentially policies, which are also created to serve a particular political purpose: and therefore the accountability for the decision made and the policy action taken on the back of it needs to rest with the human that designed it.

People can be inadvertently unfairly discriminated by some data science techniques (e.g. machine learning). This could be by excluding certain groups from the data, using proxy data to discriminate on race, or more predictively profiling certain groups). For example, public services monitored social media feeds to see the effect Hurricane Sandy was having. However, the worst hit areas were in places where fewer people had smartphones, and therefore went unnoticed[26].

On the one hand, people's collective provision of data allows organisations to target individuals based on their comparison to algorithmic identities which allow them to receive better, targeted services. On the other hand, by just providing them with information or services suited to their needs excludes them from ever receiving anything different. If "algorithmic information services can be personalised to this degree, the diversity of public knowledge and political dialogue may be undermined"[27].

Predictive techniques which provide 'reputational feedback' are powerful ways to regulate or nudge people to change their behaviour. *"But this only regulates or governs the effects of how they are acting and not the underlying social injustices that cause the effect. The devil doesn't wear data. Social injustices are much harder to track than the everyday lives of the individuals whose lives they affect."*[28]

[23] Anderson, C in Crawford, K (2013) The Hidden biases in big data

[24] Crawford, K (2013) The Hidden biases in big data

[25] Gillespie T (forthcoming) The relevance of predictive techniques , forthcoming in Media Technologies (eds) Tarleton Gillespie, Pablo Boczkowski, Kirsten Foot, MIT Press

[26] ibid

[27] Sustein in Gillespie

[28] Ibid

Public dialogue could give policy makers an insight into the public's outlook on the challenges presented by the greater use of predictive techniques, in order to understand how to tackle them.

f. Transparency

Key questions for public dialogue:

- What level of transparency does the public expect from government on their use of data?
- What are the dangers of transparency?
    - For example, how transparent should predictive techniques be? Transparency could make the system vulnerable to gaming or abuse
- What are the dangers of a lack of transparency?
    - For example, a lack of transparency can lead to discrimination or bias in predictive techniques going unchecked

Since one of the core areas of public concern is around not knowing how, why and in what way their data is being used, which can lead to suspicion of government and a lack of trust, transparency is an important part of the data debate. Clarity and transparency on what policy makers are not doing is also key to addressing the public's concerns. Public dialogue can help policy makers to understand the level of transparency the public expects from government on data issues, and their view on the dangers both of transparency and a lack of transparency.

Transparency of process can highlight good practice, demonstrating that government is acting both ethically and for social benefit. There are significant positives if government can demonstrate that it is transparent about: the data it is and isn't collecting, how and why it is collecting data, what it is doing with the data, how it plans to share data, how data is stored and how it is ensuring security. It builds trust with the public and puts their mind at ease on what the government's principles and intentions are. For example, there are many who argue that predictive techniques, particularly if they inform public policy and the way Government makes decisions about entitlements and disentitlements, should be open and understandable in the way that legal codes and policy are. Transparency can also shine a light on inappropriate behaviour, for example, if data science projects use machine learning techniques which have inadvertent bias or discriminative built into them.

It is not enough just to put information about what government is doing out there; true transparency requires information to be accessible, understandable and 'assess'ible, and trust can be further heightened by interactive communication about the information.

There are some challenges to achieving transparency and accountability, for example, it can be difficult for non-experts to understand the algorithms – because it is genuinely complex or because it is written in a computational language not understood outside the world of data science, or because they have morphed in shape due to the feedback of the data they process.

There can be a limit to how transparent it is sensible for the government to be with its use of data. Government can be clear about policy outcomes, but there is a balance to be made between openness and transparency and allowing people to see enough information to be able to manipulate government systems and procedures. For example, if the government is transparent about the algorithms it uses to detect wrong-doing, for example, sham marriages, this can leave the system open to gaming and hamper the government's ability to detect this wrong-doing.

Public dialogue could play a role in helping policy makers to assess the level of transparency the public wants or expects from government on data issues, and understand a way to mitigate the dangers transparency, or a lack of transparency, can present to government.

5 Conclusion

The greater use of data and data science is an exciting and developing area. The technology is advancing rapidly, and government is starting to pilot different ways of using data to improve policy. However, policy makers are aware that it's a sensitive area for much of the public and civil society, with concerns around privacy and data misuse being emphasised in the public dialogue exercises carried out so far.

Policy makers understand that the public voice needs to be heard alongside others as part of the policy process. This thought leadership piece looks at the benefits of bringing in the public voice, lays out some of the key questions that the public voice can help with, and outlines the findings of some of the public dialogue work carried out to date.

It also identifies some of the ways in which the public voice on data is of use to policy makers in a broader sense:

- Greater transparency and legitimacy
- Bringing in the public voice can increase trust and avoid the public rejection of policy.
- Citizens can help policy makers to understand the ethical issues around the greater use of data.
- It is difficult for policy makers to act with confidence on complex and controversial issues when they are unclear on the public's outlook on the issue.
- Bringing in the public voice can help policy makers to identify questions or issues they hadn't thought of.

In terms of taking the next steps with public dialogue, policy makers can make use of Sciencewise as a resource to help think through the options on the public dialogue tools available to them. Policy makers can find information on public dialogue on the Sciencewise website.

The greater use of data and data science could lead to more responsive, effective and efficient public services. But first and foremost we have to have a conversation with citizens about how we can make the most of these opportunities in a responsible and ethical way.

Upcoming studies

The Nuffield Council on Bioethics report on Biological and Health Data is a major analysis of the ethical issues relating to developments that facilitate the collection, linking, use and exploitation of data relating to individual people have become increasingly important to biomedical research, healthcare, and other aspects of contemporary life. The report is due to be published in the next few months. See more at: http://nuffieldbioethics.org/project/biological-health-data/#sthash.jUNcAIBg.dpuf

Bibliography

Abelson, J & Gauvin, F-P. (2006) Assessing the Impacts of Public Participation: Concepts, Evidence and Policy Implications. Ottawa: Canadian Policy Research Networks

Anderson, C in Crawford, K (2013) The Hidden biases in big data

Bateman, Sue (2014) Open Policy Making, 'The Data Science in Government programme: using data in new ways to improve what government does' https://openpolicy.blog.gov.uk/2014/05/30/data-science-1/

Crawford, K (2013) The Hidden biases in big data

Data Science Research Institute (2014) Event Summary: The Social, Cultural, & Ethical Dimensions of "Big Data"

Gillespie T (forthcoming) The relevance of predictive techniques , forthcoming in Media Technologies (eds) Tarleton Gillespie, Pablo Boczkowski, Kirsten Foot, MIT Press

Greenhlagh, Madeleine (2014) Open Policy Making, 'The Data Science in Government programme: progress so far' https://openpolicy.blog.gov.uk/2014/09/18/data-science-2/

Howard, A (2014) data-driven policy and commerce requires algorithmic transparency

Hull University Business School (2014)

IBM (2014) 'Big Data at the Speed of Business' http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html

Institute for Insight in the Public Services (2008) 'Data and Privacy: How concerned are citizens about data sharing in the public services?

Ipsos MORI (2014) Dialogue on data: Exploring the public's views on using administrative data for research purposes

Ipsos MORI (2014) Public attitudes to science 2014. 'Attitudes to Big Data' section

OECD, 2009, Focus on Citizens: Public Engagement for Better Policy and Services

Open Data Institute (2014) 'What makes data open?' http://theodi.org/guides/what-open-data

Open Government Guide: Engage citizens in deliberation on a priority issue http://www.opengovguide.com/commitments/engage-citizens-in-deliberation-on-a-priority-issue/

The Royal Society (2012) Science as an Open Enterprise

Sciencewise, Social Intelligence: Big Data (2014) http://www.sciencewise-erc.org.uk/cms/assets/Uploads/SocialIntelligenceBigData.pdf

The UC Berkeley School of Information, 'What is data science?' http://datascience.berkeley.edu/about/what-is-data-science/