Open data dialogue
Final Report

TNS BMRB

Darren Bhattachary and Douglas Dalziel

Contents

1. Executive summary

Introduction

Undertaken on the behalf of the Research Councils UK in partnership with JISC, the Royal Society and Sciencewise-ERC[1], this public dialogue explored views on open data, data reuse and data management policies within research.

The public dialogue was designed to:

- Provide insight on the business issues that the dialogue will support, at the research councils and JISC
- Build on prior work in the area and account for the wider policy framework
- Engage people meaningfully around this complex area, enabling the public to frame issues and test out any principles emerging across a range of research contexts.

The research comprised a number of elements:

- an initial literature and policy review of the area
- two reconvened discussion groups in Swindon and Oldham
- a workshop involving key stakeholders conducted between the first and second wave of the public dialogues.

A total of 40 public participants were recruited - demographically reflecting the gender, socio-economic group, age group and ethnicity profiles of their local area.

Key findings

1. The principal arguments for open data – around catalysing innovation and building trust – were contested.

The public had mixed views on whether open data practices would build trust and promote innovation. Whilst openness was believed to promote scrutiny which could help build trust, participants were concerned that confusion may arise from multiple interpretations of the same data, which in turn could impact on the trustworthiness of research. There was no clear relationship between trust and the provenance of data (for instance trusting a researcher's interpretation of a dataset more than a commentator).

Regarding innovation, the arguments around better efficiencies in the research process, potential cost savings and to a lesser extent growth (by utilising datasets to develop new products and services) was accepted. However, the

---

[1] The Sciencewise-Expert Resource Centre (ERC) is the UK's national centre for public dialogue in policy making involving science and technology issues. See www.sciencewise-erc.org.uk

concept of openness sat uneasily with the ability for researchers to exploit their own data. There was a strong view that those who had put the effort into developing a dataset should have a period of time to take exclusive advantage of this.

2. Open data was an abstract issue for participants to engage with

Overall, participants found the concept of open data to be abstract and relatively hard to engage with. There was particular confusion around the distinctions between data, information and knowledge. Moreover, the principal benefits of open data were seen to accrue for researchers rather than the public. Participants did not express an interest in personally exploring datasets – rather they were more interested in the results and implications of research.

3. Main concerns around open data relate to promoting the public interest and protecting personal data

The most important concern around open data was that it should be promoted when it serves the *public interest.* This was *defined almost exclusively in terms of data that can help improve human health and, to a lesser extent, the environment*. It was also viewed in terms of not impacting on national security. In addition to this substantive definition of public interest, the public were concerned that *data should not be released too early or in a way that would be likely to promote poor decision making or do harm.*

Consent and in particular confidentiality around personal data were highlighted as prominent public concerns. Whilst generally, with appropriate governance provisions, participants were relaxed around data confidentiality, a significant minority were very concerned about this issue. A number of points were raised in this context:

- consent for data reuse may be conditional
- combining data may reduce confidentiality
- consent for reuse may lead to data being used for purposes or by companies which individuals do not agree with
- lack of clarity around the ownership of linked data sets[2] could create problems around data use and consent.

Costs around the data curation and time spent by researchers enabling datasets to be prepared that could be otherwise be spent on research, or other public policy issues also were of concern to participants.

---

[2] This refers to data from a variety of sources, such as research data and administrative data (for example crime data, health data, etc) that can be combined into larger data sets using shared datafields to link the data.

4. **Current practices around open data seen as mixed and contingent on the interests of researchers**

Whilst public funded and academic researchers were generally thought to be more open than those funded in the private sector, all practices around data openness were seen to relate to the interests of the individual and the organisational context. Research culture was seen as particularly important in this regard; with some areas of research viewed as more open than others. Increased commercial funding within universities was seen as having the potential to negatively impact on academic culture, which was viewed as relatively open.

5. **The public identified 8 key principles that could be used to promote more effective open data policies. Defining the public interest was key in this context (see point 2 above).**

    i. Publically funded data should be open, unless not in the public interest
    ii. With private or co-funded data there was a right not to disclose, unless in the public interest
    iii. Researchers should be allowed a short period of exclusive access to data to create value
    iv. Personal data should be confidential, and consent gained for future use
    v. Anyone reusing public data should acknowledge the source
    vi. Data openness should be governed through an independent group
    vii. Data should be checked for inaccuracies before being made open
    viii. Raw data should include full details explaining what the data relates to, how it was collected, who collected it, and how formatted.

6. **There was a series of social and ethical issues associated with rights and responsibilities of different stakeholders around open data. These were sometimes in conflict.**

Researchers had the right to: exploit data for publications; reuse public data. They had responsibilities to: keep data in a format to enable reuse; develop data management plans; protect individuals’ data and confidentiality. In addition to these issues, private funded researchers had the right not to publish their data for commercial reasons unless in the public interest. Public funded researchers should have an active presumption of publication, within a specified time period.

The public had the right to: have informed consent for personal data use; have some overview and scrutiny of governance arrangements around personal data; have research data funded by the taxpayer reused for collective gain; have research data in the public interest disclosed (accounting for the principles in point 5).

Funders had the right to: expect data management plans to be developed and actioned. They had responsibilities to: develop appropriate governance mechanisms to scrutinise public interest; develop contracts with the private sector that seek appropriate balance in data openness; enforce and publish data management policies and practices.

7. Different governance structures emerged around open data, which are contingent on how public interest is defined and protected.

There were three broad governance arrangements outlined by participants, which need to be developed in relation to whether data is in the public interest:

- For research where data was not considered to be in the public interest: participants were content for 'self-governance', whereby other researchers and funders, who have technical knowledge and understanding, would oversee open data practices.
- For research where data had public interest implications (e.g. where it affected human health or had a significant impact on the environment), participants wanted a wider range of specialists (e.g. ethicists, lawyers, economists, Non-Governmental Organisations) to be involved in data governance.
- For research in the public interest where there were implications around personal data; in addition to other specialists, the lay public should also be involved. In certain instance (e.g. medical datasets) regulations should also govern this area.

Publishing data management plans was endorsed across the board. Finally, it should be noted that certain areas of research were considered to be of low public interest, and not worth the time and expense to curate datasets.

Conclusions

8. There were two broad views on what constitutes the public interest around data reuse that emerged from the dialogue, which have distinct and potentially conflicting implications for data governance.

The first was an instrumental definition and directly relates to whether data has public interest implications. Here data was believed to be in the public interest if it provides a direct utility or benefit. The second relates to a public interest definition that not releasing data when there is the potential to do harm – for instance by encouraging poor decision making. Governance in this context becomes more paternalistic, with data release contingent on the quality of the evidence, timing, whether research has been replicated and so on.

On balance, however, there was a presumption of active publication of publically funded or public interest data. As such the first definition should be the driving principle behind data openness, whilst mindful of negative consequences of publication.

9. Opening data in the public interest can involve direct trade offs for the public – in terms of individual rights around consent and confidentiality and the greater good.

While a small but significant number of participants were keen to gain consent each time personal data was reused, this may be very difficult to achieve in practice given complexity of data linkage, ownership and wider governance.

Related to this, greater good arguments – particularly around public health outcomes – were seen to override individual rights in many contexts.

Given the potential to exploit this, dealing with it requires integrity from those overseeing the governance of data not to overstate claims around research, particularly on controversial or contested areas, and to appropriately consider rights of participants as well as potential beneficiaries.

10. There are a number of implications emerging from the dialogue about who gets to define public interest.

In part this will be defined by disciplinary area, with certain avenues of research - most obviously medical - likely to be by definition in the public interest.

Beyond this it is worthwhile thinking about how the current administration of research grants can be used when thinking about data governance. These existing processes could be used to help categorise research into areas with higher public interest, which in turn may need distinct practices and structures around data governance – namely accounting for public principles around openness and the participation of lay people.

11. There is an opportunity to directly build public principles around openness directly into the policies of research funders.

There was a strong resonance between views of the public and views on openness emerging from the Research Councils, with findings in particular complementing RCUK Common Principles on Data Policy. A focus could now be on improving, codifying and reviewing practices to ensure these aspirations are met. There is a significant gap between such views on openness and current practices.

12. Different open data governance structures could be considered by the research councils.

For research data with high public interest (such as relating to impacts on human health and the environment, or national security; or where personal data involved) lay or wider professional involvement should be considered.

Research data with lower public interest (defined as not relating to human health, security and environmental impacts) were believed to be adequately overviewed by researchers and the research councils

Data management plans should be published and policed, with sanctions potentially imposed on researchers for non-compliance.

13. Open data, though not unimportant, is unlikely to be a highly significant public issue around the governance of research

Open data did not capture participants' imagination and overall the issue was considered of greater relevance for researchers than public *per se*. Given also the complex relationship between openness and trustworthiness, addressing open data alone is unlikely to have a major impact on governance concerns around research - which generally centre on things such as the motivations of researchers, who is funding the research, the speed of research relative to regulatory capacity, the handling of uncertainty and so on.[3]

If the goal is to help promote trustworthiness, using the debate around open data to promote reflection by institutions on other areas for open governance would be worthwhile.

---

[3] Chilvers, J. & Macnaghten, P. (2011). The Future of Science Governance: A review of public concerns, governance and institutional response.

2. Introduction

2.1 Background to the dialogue

Open data is an increasingly important issue in research and policy development. A key theme of the coalition government, open data is viewed as helping to promote innovation and growth by catalysing new markets and innovative products and services, as well as improving standards and transparency in public services.[4] In this context, the Cabinet Office has encouraged a wide range of public datasets to be made openly available and established a Public Sector Transparency Board and Transparency Principles to help embed openness in practice.

Advances in technology are also increasing the opportunities to reuse and combine data sets to create new and innovative information services and products – both commercial and non-commercial. Increases in computing power, together with improvements in the data sharing infrastructure, promote the potential for large scale analysis of datasets.

However, the issue of open data in research is complex. Controversy around UEA climate data emails sparked a major debate around access to scientific data, which led to Sir Muir Russell arguing for improved quality assurance and accountability.[5] In this context, improved transparency around research is believed to help promote public trust.

Nevertheless, there are legal, ethical and commercial constraints on the release of research data: not least intellectual property rights, issues of disclosure and personal privacy; and concerns that the research process is not damaged by inappropriate release of data.

This dialogue is concerned with how open data principles and policies around research should be developed and practiced. Specifically, when considering open data:

- Should the rules of engagement be the same for publicly funded science as for private?
- What if private research has public implications?
- How do developments in new media shape this landscape?

---

[4] HM Treasury (2011). Autumn Statement. p. 40. Available at: http://cdn.hm-treasury.gov.uk/autumn_statement.pdf

[5] The Independent Climate change E-mails Review- Chair Sir Muir Russell. Available at: http://www.cce-review.org/pdf/FINAL%20REPORT.pdf

- How are public and private interests protected and promoted?
- And what is the attendant impact on public trust?

The dialogue was undertaken on behalf of the Research Councils UK in partnership with JISC. It was part funded through Sciencewise-ERC,[6] and also informing the Royal Society's Science as a Public Enterprise (SAPE) working group on access to research, the dialogue was specifically developed to provide public insights into future open data, data reuse and data management policies across the research councils and beyond.

2.2 Objectives

The objectives of the study[7] were as follows:

- to work with major UK research funders to provide public insight and feedback on future data openness, data reuse and data management policy options
- to engage key policy stakeholders to clarify the range of issues and options in policy going forward, related to the SAPE conclusions
- to build on previous studies and work alongside relevant working groups/task forces recently established by government, to help shape future research data policy within the UK's main funding institutions
- to explore wider ethical and moral issue related to open data and data reuse
- to engage business on some of the issues relevant to privately and joint funded research

Our approach to explore these issues is described in the next chapter.

---

[6] The Sciencewise Expert Resource Centre (-ERC) funded by the Department for Business, Innovation and Skills (BIS), aims to improve policy making involving science and technology across Government by increasing the effectiveness with which public dialogue is used, and encouraging its wider use where appropriate to ensure public views are considered as part of the evidence base. See: www.sciencewise-erc.org.uk

[7] Note: The objective: "to formulate views on the policy conclusions from the 'Science as a Public Enterprise' (SAPE) Royal Society working group study" which was part of the original brief was not included in the final specification of the research - due to timetable changes for the SAPE report. However, a close synergy was maintained between both projects through Royal Society representation on the Open Data Dialogue Oversight Group.

3. Our approach

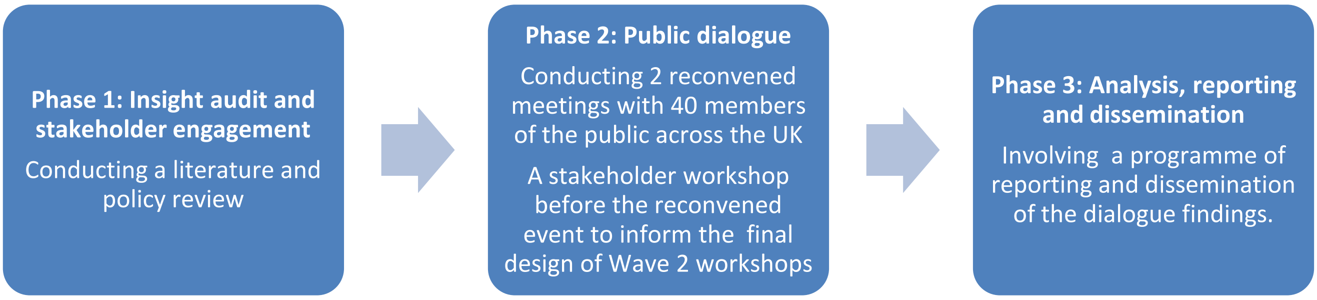
In this section, we provide a summary of the method. Fuller details of the method are provided at Appendix 1.

3.1 Overview

This is an important study for the research councils and a wide range of influential science, policy and business stakeholders who will be interested in the outcomes of the research. The public dialogue was designed to:

- Provide insight on the business issues that the dialogue will support, at the research councils and JISC
- Build on prior work in the area and account for the wider policy framework
- Engage people meaningfully around this complex area, enable the public to frame issues, be responsive to their concerns, and test out any principles emerging across a range of research contexts

To address this, we developed the following 3 stage approach:



3.1.1 Phase 1: Insight audit

This phase was comprised of two components:

- a set up meeting to scope the policy context and make revisions to the approach
- an information review to explore current principles and policies on open data, around as well as views on the benefits and concerns around changing these practices.

3.1.2 Phase 2: Public dialogue workshops and stakeholder engagement

This phase comprised of three components:

- Wave 1 of the public dialogue;
- Stakeholder mapping and scoping workshop
- Wave 2 of the public dialogue

The dialogue convened 40 members of the public and was held in two areas - Swindon and Oldham. The same participants were engaged for Wave 1 and Wave 2.

The objectives of the Wave 1 workshops were to:

- Explore top of mind views around openness and how it relates to research
- Introduce and inform people about the data openness
- Enable people to think about principles governing data openness
- Develop a list of key principles to be explored in the next workshop

Between Wave 1 and 2 workshops people were asked to research the issues and speak to friends and family about open data. This enabled us to reach a wider sample of people indirectly, and provide insight into impact of external information on views.

Between Wave 1 and 2 a stakeholder mapping and engagement exercise was undertaken to understand stakeholder needs from the dialogue and gain feedback on the materials and approach for the dialogue. The stakeholder workshop was held on 23rd February 2012 at the offices of TNS BMRB, in London.

The objectives of Wave 2 workshops were to:

- Answer any questions about open data arising since Workshop 1
- Understand and discuss how open data works in practice in different settings - by hearing from private and publicly-funded researchers
- Explore different perspectives on open data - using actors to bring to life case studies
- Review the principles governing open data
- Consider who need to be involved in thinking about the public interest in open data policies

Details of the sampling and recruitment approach are provided at Appendix 1.

1.1.1 Analysis, reporting and dissemination

The workshop discussions were digitally recorded and transcribed before an analysis of the data undertaken. The analysis consisted of a combination of content and framework analysis. At the time of writing, dissemination activities are being planned. As part of this, emerging findings have been fed back to the Administrative Data Taskforce.[8]

---

[8] The Administrative Data Taskforce, established in December 2011, seeks to propose new mechanisms and collaborative agreements to enable and promote the wider use of administrative data for research and policy purposes. It is tasked to report to Ministers within 12 months. See: http://www.esrc.ac.uk/funding-and-guidance/collaboration/collaborative-initiatives/Administrative-Data-Taskforce.aspx

4. Findings

4.1 Initial views on openness and research

When considering their personal lives, participants recognised a continuum between the benefits and drawbacks of being open. Whilst being open can promote trust and understanding, it can also detract from both under certain circumstances.

When considering research data, People already feel that they are overloaded with information and opening up data could increase disputes around interpretation.

People did not necessarily make a link between provenance, the interpretation of data and trustworthiness.

Participants were asked their views about how open researchers were. A number of key issues emerged. While there is recognition that amongst researchers there is a culture of sharing data to advance knowledge, researchers varied in the extent to which they were open. Those working in the public sector were considered to be more open than those working for private companies. Openness was seen as constrained by the researcher's personal and professional interests.

4.1.1 Being open

Participants were first asked about what they understood by "being open in their own lives" and what benefits that brings. Participants had both positive and negative associations of what being open meant to them.

4.1.2 Positive associations of being open

Participants tended to think about the idea of being open in terms of their relationships with family, friends and acquaintances. Their initial top of mind associations tended to be largely positive. Being open was seen as a good thing, as it:

- Builds trust
- Promotes better understanding
- Allows evaluation and helps improve decision-making
- Encourages reciprocal understanding

Amongst friends and family being open was considered to promote engagement and strengthen friendships.

> "I think if you are not an open person you come across as unapproachable – being cold."
> Oldham, Wave 1, Group 2, Female.

Participants considered that disclosing information about themselves and their feelings in a safe environment with friends and families demonstrated trust and in turn this would be reciprocated. For the listener, sharing information confirmed feelings of being valued. Where problems were discussed with friends or family it was felt that being open increases understanding for both the person being open and the listener.

Participants also considered that being open and sharing information was useful to help reach better decisions. In a friendship, sharing information invites reflection on the information, evaluation of the information and improved decisions or choices.

> "You have more information about something, it informs you so that you can make your own choice, and you can make an educated choice then."
> Swindon, Wave 1, Group 1, Female.

In professional relationships too, it was recognised that sometimes it was essential to be open and honest to enable people to come to a fully informed decision. For example, being open with medical practitioners was seen as important to help doctors come to an accurate diagnosis and to decide the best form of treatment.

> "Yes when you go to the doctors, for instance, you have to be very honest with the doctor, haven't you? Telling him what the problem is. Otherwise, you are not going to get the help that you need."
> Oldham, Wave 1, Group 2, Female.

4.1.3 Negative associations of being open

Participants however quickly recognised that that sometimes being fully open and sharing information was not always a good thing, as:

- Too much information can confuse or disempower people
- It can break trust
- Some things are considered private

- In some instances sharing can harm people

Participants described how it was possible to provide “too much information”. This led to three difficulties.

Firstly, it led to confusion. Sometimes providing information could have unexpected negative results:

> “I have an elderly mother and they tell her the side effects of a certain tablet could make you do this or that and instantly she has got these side effects, so sometimes the information is not necessary for some people. What I am saying is the information is good if you can take the information.”
> Swindon, Wave 1, Group 1, Female.

Secondly, it led to difficulties ’adsorbing’ information. Too much information could overwhelm listeners and make it difficult to take in. When there was a lot of information from different sources, and where it was contradictory, it was considered to disempower people and make it impossible for them to make decisions or take action. Information that is difficult or impossible to do anything about can also disempower people:

> “I just think some people live with the fear of the information they have been given.”
> Swindon, Wave 1, Group 1, Female.

Thirdly, in some cases being open can break confidence and trust. Where views about other people are shared and these views are then passed on, it can affect relationships. Being open with people can make them question the things they felt confident about previously. It can make people uncertain about the validity of the information and make them question the source of the information. In some cases this can cause people to lose trust in the person providing the information.

Certain information was considered to be too private to share widely. Participants talked about Facebook photographs, which could be restricted to friends. Sharing these more widely was considered to make you vulnerable to misuse of your information. Participants were aware of reports of employers looking at this information and using it against employees or potential employees.

Participants were clear that sometimes they did not want to have certain information disclosed to them.

4.1.4 Judgements around being open

Participants believed it was unusual for people to be fully open, and considered that when people did reveal information it was important to understand their motives for being open.

> "I think people are not open at all."
> Swindon, Wave 1, Group 2, Female.
>
> "Yes I think you assess somebody within five seconds of meeting somebody basically. Almost as soon as they open their mouths."
> Swindon, Wave 1, Group 2, Male.

Participants talked about the importance of restricting information where there was a perceived self-interest. They also discussed the importance of making judgements about what should be revealed, to whom, and when, and the risks of "being open". Being open could make people vulnerable to others who might misuse their openness.

> "People can also abuse the situation if you give too much away."
> Oldham, Wave 1, Group 2, Female.

Making good judgements about how open to be was essential to make better decisions, and to avoid disempowering people. However, it required information about a number of factors which helped speakers assess the risks. The circumstances and surroundings were considered to be important in assessing the extent to which people are prepared to open up to each other.

4.1.5 How open are researchers?

Participants were then asked their views about how open researchers were. A number of key issues emerged:

- Amongst researchers there is a culture of sharing findings to advance knowledge
- Researchers varied in the extent to which they were open
- Researchers working in the public sector were considered to be more open than those working for private companies

- Rosewaters openness is constrained by their employers' interests

The general public considered that they were not very aware of what researchers did, and had to rely on media reporting to inform their views. It was felt that media reporting may not be completely objective, and would highlight newsworthy examples where researchers chose not to share information.

> "I don't think we hear enough about what scientists do and what they are researching."
> Swindon, Wave 1, Group 2, Female.

At a generic level, participants considered that accumulating knowledge was a process that relied on researchers sharing information about their work with colleagues in order to build knowledge. The culture of sharing the results of scientific work with other researchers appeared to be essential to make progress:

> "I think for the most part scientists are an open group of people who like to share their knowledge, like to expand their knowledge and like other people to understand exactly how far they have reached. Scientific knowledge has got us to where we are at the end of the day."
> Swindon, Wave 1, Group 1, Male.

Participants considered that researchers varied in the extent to which they were open. There was recognition amongst participants that researchers were sometimes in competition with other researchers to be first to make a discovery or to perfect a technique and that as a result they may not always be open about their data. There was a strong view that researchers also liked to be very confident about their results and so might not be prepared to be open until they are certain of their findings.

Participants considered that academic researchers funded by the public sector would be more likely than those working in the private sector to be open. Those working in the private sector were seen as constrained by commercial interests in what they could be open about.

> "They are not always open ...because they're being paid by somebody to do what they do."

> Oldham, Wave 1, Group 2, Male.
>
> "It depends who is sponsoring it really. I mean, if you're talking medical research and you're talking development of medicine , well you're talking say private limited companies who are privately sponsoring their own research into things, so in that respect they are not going to give you the information.
> Swindon, Wave 1, Group 2, Male.

There was considerable debate about whether researchers working in private companies were more concerned about commercial interests or the public interest. Participants felt that there were a number of accounts of companies withholding information that was in the public interest. For example, the tobacco industry was seen as not serving the wider public interest by withholding information about the link between cancer and smoking.

> "Tobacco is the classic case, and if you look a lot of the research that Imperial Tobacco had way, way, way, way back, and its only coming out recently, that they knew about – that was self-interest."
> Oldham, Wave 1, Group 2, Female.

Participants were in agreement that where public interest was at stake then data should be made open, even if it affected the commercial interests of private companies.

The recent story about PIP, the French manufacturer of breast implants that was linked to increased risk of cancer, was discussed by participants across the different groups. Participants considered that the private company should not hold back data in the investigations even if it risked the commercial viability of the company.

> "Any company or anybody that actually does stuff that affects the public like breast implants, they shouldn't be able to withhold that for private interest."
> Swindon, Wave 1, Group 1, Male.

However, assessing public interest was not seen as straight-forward. Participants expressed the view that coming to a decision about data involved considering a range of wider issues: trust in the researchers, trust in the company or university, and information about who is funding the research. This was considered to be particularly difficult to resolve when researchers expressed contradictory views.

> "Different scientists saying different things...One saying there is such a thing as global warming and the others who presumably work for the companies that have a vested interest and their scientists are saying no it doesn't."
> Oldham, Wave 1, Group 2, Male.

The debate about the MMR vaccine was discussed, and here the key conclusion was that while sometimes researchers might have a different viewpoint based on different data sets, what was required was that data is shared – not just the interpretation of the data.

Participants considered that if the data is made open then there is the opportunity for other researchers to validate the data, and make decisions about the quality of the data.

> "Now if other people could have validated his data, they might have come to a different conclusion."
> Oldham, Wave 1, Group 2, Male.

4.2 Understanding of open data

Open data was an abstract issue for participants to engage with

- Overall, participants found the concept of open data abstract and relatively hard to engage with. There was particular confusion around the distinctions between data, information and knowledge. Moreover, the principal benefits of open data were seen to accrue for researchers rather than public. Participants did not express an interest in personally exploring datasets, and were more interested in results of research.

- The key benefits of open data were considered to be: scrutiny of the data, reduction of duplication of research effort, and opportunities to reuse the data.

- The public were concerned about the mis-interpretation of data, and how open data might increase confusion over the interpretation of results.

4.2.1 Public understanding

Participants viewed a video film of researchers discussing data: how it is collected, who has access to data and how data is controlled. They were then invited to discuss the contents of the video.

4.2.2 What is data?

Participants had limited understanding of what constituted data: for the most part they tended to relate it most easily to medical examples, and particularly to the outcome of drug trials. Only a minority considered they were familiar with different types of data.

Although not universal, participants were unclear about the distinctions between data and information/knowedge and tended not to distinguish between reports, academic publications, and data. Throughout this and subsequent sessions the words “data” and “findings” were frequently confused.

However, there was an appreciation that special techniques were required to ensure that data was collected in a way that was reliable and valid. In particular there appeared to be an understanding of the need for controlled conditions and experiments.

4.2.3 Perceived benefits of open data

Respondents had mixed views on the benefits of open data. They considered that the potential benefits were likely to be:

- Opportunities for other ‘professionals’ to scrutinise the data
- Reduction of unnecessary duplication of research (as well as a saving of time and money). This was seen as a major benefit particularly amongst those who were concerned that there was not a central register of research activity
- Opportunities to build on the learning from existing data or to reuse the data in further research

However there was considerable debate about whether these benefits could realistically be achieved.

4.2.4 Scrutiny of the data

Participants were concerned about the quality of open data, in particular if it was reliable. Participants felt they would not be able to judge whether the information was reliable and considered that only another scientist would be able to judge this.

As such, there was considerable interest in ensuring that data was verified before being made open, and that professionals would need to have responsibility for doing this. Participants considered this important to check the veracity of the data and to prevent false claims being made about the findings.

Participants mentioned how the researchers at CERN had behaved in a positive way by being open about the data they had on particles travelling faster than the speed of light – which has allowed other researchers to try to validate this.

> “Now, if you made that data available to a third person and they come to a similar conclusion as you, surely that’s more beneficial to you.”
> Oldham, Wave 1, Group 2, Female.

4.2.5 Reducing duplication

Reduction of duplication of research effort was seen an important benefit of open data. It would mean that researchers would not repeat research unnecessarily and this would save both time and money. Opening data in areas where research had produced negative results was seen as important - both to focus limited resources (by learning from failed experiments) and to increase transparency.

Participants envisaged that the mechanisms for reducing duplication might lead to a centralised catalogue of ongoing research, rather than the current arrangements where funders are mainly responsible for keeping information about previous and current research.

> “There should be some overarching body that’s keeping an eye on all of the research that’s going on. I don’t think scientists should be allowed to operate in complete isolation.”
> Oldham, Wave 1, Group 2, Male.

Participants mentioned potential cross-fertilisation benefits from keeping open data centrally: as research covers such a vast area, it might mean that researchers in one area would have easier access to findings from different areas they did not normally look at, which may be beneficial to them.

Opportunities to build on the learning from others’ data and to reuse data were seen as a significant benefit by respondents. Participants considered that researchers with limited funding would be able to examine data that had been produced by those with more significant funding, and that this allows the smaller players to contribute intellectually to the research. It was felt that other researchers might reuse the data in different ways, or link it with other data sources to advance thinking and develop new ideas.

4.2.6 Interpretation of data

Participants did engage with the discussions about how data is interpreted to draw reliable conclusions. Participants were concerned about how researchers could interpret data in different ways and the impact of misinterpreting data.

Discussion of the MMR vaccine brought out a number of wider issues. Respondents were concerned about the personal motivations of researchers who were 'out to make a name for themselves'. Publicity-hungry researchers who put personal ambition before scientific rigour were seen very negatively.

> "He was gaining kudos and glory from it, wasn't he? Which is always going to a problem with scientists, I think."
> Oldham, Wave 1, Group 2, Male.

Participants voiced concerns about how the media promoted changing and contradictory views about the benefits of diet and health regimes which different researchers espoused. Reports about the benefits of eating certain foodstuffs or partaking in certain health regimes which later turned out to be false or misleading were a particular issue that engaged participants. This reinforced the views that researchers may be motivated by self-interest and that data needed to be independently verified before findings were acted on.

Participants considered that poor interpretation of data could lead to mistrust of the scientific community:

> "I was going to say, you hear all these different things and you end up making your own decisions, you just think oh well, I won't take much notice of the data."
> Swindon, Wave 1, Group 1, Female.

Participants were asked about the provenance of data, but were of the view that they did not trust one source over another one.

4.3 Views on Open Data Principles

The public identified 8 key principles that could be used to promote more effective open data policies. Defining the public interest was key in this context

1. Publically funded data should be open, unless not in the public interest
2. With private or co-funded data there was a right not to disclose, unless in the public interest
3. Researchers should have a short period of exclusive access to data time to build value
4. Personal data should be confidential, and consent gained for future use
5. Anyone reusing public data should acknowledge the source
6. Open data should be governed through an independent group
7. Data should be checked for inaccuracies before being made open
8. Raw data should include full details explaining what the data relates to, how it was collected, who collected it, and how formatted.

4.3.1 Understanding of principles

Towards the end of the first wave of the public dialogue, respondents were asked to reflect on their earlier discussions and begin to consider what principles should guide RCUK in their thinking about open data.

Initially this discussion involved an open exchange of ideas. Later in the discussion a list of possible principles was presented to respondents on cards. The principles included some that respondents had discussed (on blue cards) and others that had been developed by the Public Sector Transparency Board and RCUK for this research (on yellow cards). Respondents were asked to sort these into high medium and low priority.

The public dialogue revealed that 8 key principles were consistently ranked as high priority across the groups. These were as follows:

- Publically funded data should be open, unless not in the public interest
- With private or co-funded data there was a right not to disclose, unless in the public interest
- Researchers should have a short period of exclusive access to data to build value
- Personal data should be confidential, and consent gained for future use
- Anyone reusing public data should acknowledge the source

- Open data should be governed through an independent group
- Data should be checked for inaccuracies before being made open
- Raw data should include full details explaining what the data relates to, how it was collected, who collected it, and how formatted.

The **public interest** was defined almost exclusively in terms of data that can help improve human health and, to a lesser extent, the environment. It was also viewed in terms of not impacting on national security. Data release was also not in the public interest if likely to promote poor decision making or do harm.

4.3.2 Publically funded data

There was an almost universal agreement that publically funded data should be open, unless it is not in the public interest (e.g. due to national security concerns). Respondents considered that as this data was funded by the taxpayer the public had a right to access this data. Respondents spoke in terms of the public "owning" the data and wanting to ensure that it provides the maximum return on investment. Sharing the data, particularly where it related to health, finance, crime and science was seen as a mechanism for making the data produce as much value as possible.

> "We are the ones paying for it…because if you're spending the money, you will see what the results are or what is actually happening behind the scenes."
> Oldham, Wave 1, Group 1, Male.

Respondents also talked about making the data open to ensure that researchers were accountable to the public. Accountability might include ensuring that the total research budget is spent well, and the research addresses areas of interest to the public. On the flipside, making the data open would ensure that the private interests of researchers are not being sponsored from the public purse. The public would be able to scrutinise the research coverage, focus and to some extent the outcomes of the research. Respondents wanted to see public data made easily accessible, by being published on public websites, with details of how the data was collected. They also wanted to see that the data was in a format that was widely compatible and accessible.

> "Public data will be widely compatible and should not require specialist equipment or expensive software to access it."

> Oldham, Wave 1, Group 2, Male.

Public data would need to be rigorous and up to date. Respondents also wanted to ensure that the data had been validated.

Respondents agreed that in some cases it may not be in the public interest to have publicly funded data made open. Examples included: national security, and other data that might be used for nefarious purposes (such as crime data).

4.3.3 Privately funded data

Respondents considered that with private or co-funded data there was a right not to disclose, unless it is in the public interest.

Participants considered that researchers and companies should have the right to seek Intellectual Property Rights (IPR) to research data. There was widespread agreement that this was acceptable:

> "I think perhaps if you have all the data and you are making a breakthrough in something, perhaps you have a commercial interest because you are funded by a pharmaceutical company. Then you are going to want your data. Whereas if you share that data too soon, other companies might come on to the bandwagon and you have lost the potential to have a worldwide market."
> Swindon, Wave 1, Group 1, Female.

> "I don't think that you can impose on a private company that they will share their data with us".
> Swindon, Wave 1, Group 1, Male.

Respondents continued to return to medical issues and to the development of drugs when thinking about this principle. Where drug trials had led to drugs that were not subsequently taken to market, the overriding view was that the company had every right not to disclose this information. Only when they released drugs did they have an obligation to provide data on the drug. However, there was strong support amongst participants that there was a need to protect the company's Intellectual Property through licensing arrangements, such as not allowing other companies to manufacture the drug for a period of time.

In addition participants were satisfied that national government agencies would need to oversee the trials and authorise the drug. However, there were some concerns that the system could not be foolproof, and that occasionally medicines such as thalidomide or devices such as breast implants might do harm.

A substantial caveat was that participants felt that where the data was in the public interest, it should be made open. They considered that examples might include where there are health benefits for the population these should be disclosed. In addition where companies had caused harm such as in the case of the breast implant company PIP, the data should be made open.

> “Unless national security is involved or something... I am suggesting that if you paid for some research to be done that is your research, but if you happen to find something on the way that was of national importance then it is taken out of your hands, so that’s the caveat.”
> Swindon, Wave 1, Group 1, Male.

Views about privately-funded research not being open were extended to a global context, in which UK private companies should not disclose data that would harm the UK economy, except in cases of wider public interest.

Health and security were the areas that participants considered were most likely to raise public interest issues.

4.3.4 Exclusive access

Participants considered that researchers should have access for a limited time period to their own data and not be under any obligation to make the data open until the end of this period. There were two considerations. Firstly, to ensure that the researcher had every opportunity to ensure the accuracy of the data, prior to making it open. Secondly, to allow researchers exclusive access so that they were able to benefit from the work they had undertaken to produce the data, through publications and so on.

Respondents were keen to ensure that there was also a time limit on privately funded data that companies keep secret which is in the public interest (such as research funded by the tobacco industry).

4.3.5 Personal data

Respondents were strongly of the view that personal data should be kept confidential and that consent should be gained for future use of the data.

> "You do have to consent for it to be shared."
> Oldham, Wave 1, Group 2, Female.

Respondents considered that it was essential to ensure that it was not possible to link open data to individuals. Respondents were concerned that failure to do this could lead to serious consequences for individuals and would open them to potential harm. This might include loss of privacy; and targeting of individuals by criminal or commercial interests.

Respondents also considered that consent to participate in a research study, did not imply that the data could be re-used without again securing the consent of the individual. This was considered to be less clear cut when data was anonymised. However research that combined data sets might allow individuals to be identified.

4.3.6 Acknowledging the source

Respondents considered that the source of data should be acknowledged, although they were aware that policing this might be difficult or impossible. However, respondents felt that this should be done in a similar way in which journal articles acknowledged other researchers contributions.

> "Well they are saying that you acknowledge the source, like someone writing a factual book, he would usually acknowledge all his sources."
> Swindon, Wave 1, Group 1, Male.

4.3.7 Governance

Respondents thought that open data should be governed through an independent group. Governance was required to develop and ensure that appropriate open data policies were effectively managed.

In addition respondents wanted to see governance arrangements that were independent. Independence required that those directly involved with the research or with interests in the outcome of the research should not be solely responsible for compliance with open data policies.

There was debate about who might be included in the governance structure and at this stage of the public debate it was felt that research bodies, researchers, professionals such as lawyers and accountants, as well as the general public may all play a role. There was a view that the arrangements should not include politicians.

> "I'm just saying that perhaps it shouldn't have been the government that makes these decisions; it should be an independent."
> Swindon, Wave 1, Group 2, Female.

4.3.8 Data checking

Respondents considered that data should be checked for inaccuracies before being made open. A major consideration was that scientists needed to be allowed the time to do this. However, there was considerable support for including a means of validating the data independently using another scientist from an independent organisation.

> "It would mean nothing, the raw data, but the conclusion that you came to would be more beneficial if a third person looked at your data and said, yes, I agree with that."
> Oldham, Wave 1, Group 1, Male.

4.3.9 Data preparation

Respondents considered that raw data should include details explaining what the data relates to, how it was collected, who collected it, and how it is formatted.

Although there was concern at the potential costs of doing this (both financial and as a result of taking researchers away from actually doing their research) it was felt on balance it was an essential requirement if others were to effectively use the data.

4.3.10 Other principles

Whilst the above 8 principles appeared to have the widest support there were other principles that respondents considered would be of lesser priority. These included principles regarding:

- The detailed administration and governance arrangements such as: a single point of access for public data, maintenance of inventories of data holding, and that data management policies should be developed in accordance with standards and best practice.
- Licensing arrangements, which respondents appeared to have little knowledge of.

4.4 Reviewing the importance of open data at workshop 2

Between the first and second workshops participants were encouraged to reflect on open data and consider wider views from family, friends and other sources including the media. Two issues rose in prominence:

- A significant minority were concerned about data being “sold on” and combined with administrative data, which may affect confidentiality.
- There was a concern that data could be rushed out without appropriate checks and this could lead to data being either misinterpreted or wilfully misused.

Between the first and second workshops, participants had the opportunity to reflect on the initial dialogue, discuss open data with friends and family, and consider issues (e.g. news stories and other media coverage) where being open around research data may or may not have been helpful from their perspective. Two issues emerged more prominently: confidentiality and misinterpretation or misuse of data.

4.4.1 Confidentiality

Firstly, the issue of personal data and confidentiality emerged more strongly for participants. In particular, there was a greater emphasis on the increasingly ubiquitous way in which personal data was held: from store card information to digital footprints.

> ‘I mean, most of our sort of dealings with data, realistically, is going to the shop or going to the petrol station and paying on the card and this data, and people know exactly what you’ve bought at what time. Realistically, that’s how people get our data. Banks getting our data, giving our data to other companies. Trust? I wouldn’t trust a banker’.
> Oldham, Wave 2, Group 2. Male

A significant minority of participants were extremely concerned around the selling of personal data for purposes of limited public benefit, which in turn impacted on trust in organisations. Bringing this together with administrative data and other research data, could be powerful and impact on the privacy of individuals.[9]

Concerns also focused on who could gain access to medical information, the purposes and motivations of those wanting this data, and the public outcomes from such research. This issue, which is explored in more depth later, has significant implications for informed consent.

4.4.2 Misinterpretation and misuse of data

Secondly, there was a significant concern that researchers would use data beyond its original intended purposes, and hence come to misleading interpretations – which could have serious implications. In this context, people were also concerned about researchers 'cutting corners', rushing to market, or not going through all the appropriate checks and balances to ensure research was robust. Related to this was the potential for 'wilful misinterpretation' of data for political or other purposes: crime datasets and opinion polls were highlighted in this instance, but also issues that could affect the interests of business - such as data relating to environmental problems or legislation. In both of these contexts, participants wished to see effective governance in terms of: being clear as to the scope and limitations of the dataset; what stage the research was at; as well as what is appropriate in terms of interpretation of the data.

4.5 Reflections on open data practices

Participants had the opportunity to listen to and discuss current open data practices with researchers and others involved in developing policies around open data in the public and private sectors. Three issues emerged:

- The benefits of open data were considered opaque, but people keep an open mind
- Though costs of open data were more easily envisaged, they did not outweigh benefits
- In principle data should be opened to enable as much access as possible, without entailing excessive costs.

During the dialogue, participants had the opportunity to listen to and discuss issues with researchers and others involved in developing policies around open data in the public and private sectors. This session specifically reflected on current open data practices, and the perceived benefits and drawbacks of

---

[9] These issues were not explored further due to the focus on research data the dialogue and the fact public views on administrative data were being explored by the Administrative Data Taskforce.

making data increasingly open. Three issues emerged which are discussed below.

4.5.1 Benefits of open data opaque, but people keep an open mind

Overall, the economic benefits of open data were not immediately clear to participants, beyond a few fairly obvious examples relating to medical innovations – such as sharing data on the efficacy of new drugs or treatments.

Certain examples were discussed in the groups, such as research suggesting that the benefit of public data assets to the European economy, if used effectively to drive innovation and enterprise, could be up to 250bn Euros.[10] Illustrations of innovations were also cited such as the release of NASA data enabling the development of desktop geographical-browsers such as Google Earth.[11] However, these illustrations did not strongly resonate with participants and overall it was difficult for people to see the perceived value in opening data for them. It may be the case that as innovation increases from such practices, so the relevance of open data will become more immediate for people.

Of greater relevance for participants were potential cost savings due to the need to build on rather than collect primary data. The ability to undertake more thorough analysis of secondary data was also recognised in this regard.

The idea of open data encouraging people to become 'citizen scientists' - looking at the data for themselves, and providing the public with a better idea of what questions researchers are trying to answer, was not seen as relevant for participants – due to a lack of appropriate knowledge and understanding to interpret datasets.

As noted earlier, there was an ambivalent relationship between data openness and trust. While openness in itself would not necessarily promote trust, participants stated they would be suspicious of researchers who were unwilling to let others scrutinise their data. A lack of openness could hence increase mistrust.

In this context, data was generally not the most significant area where participants wished for openness. It related more to funding, why one project was supported and another was not and the motivations behind this.

---

[10] McKinsey (2011). Big data: The next frontier for innovation, competition, and productivity. Available at: http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

[11] See: http://www.nasa.gov/open/plan/science-data-access.html

4.5.2 Though costs of open data were more easily envisaged, they did not outweigh benefits

The relative costs of implementing open data policies were more intuitive for participants to grasp. Estimates of between 2-5% of the annual research budget were cited in the workshop session, which would equate to £20-50 million pounds per year across the Research Councils. This was viewed as a very significant sum, and there was some debate as to whether this was best spent on data curation or more pressing public policy priorities. There was also concern as to how long data would need to be stored for.

Nonetheless, overall arguments around the greater good, together with the proposed wealth and job creation through opening data were seen to offset this – and on balance participants (within reason) felt that the costs of administration should not restrict open data polices.

A further concern related to the prospect that researchers will be required to spend valuable time curating datasets rather than getting on with their research. In this context, the idea of developing a new generation of skilled digital librarians who understand research, and also have the skills to preserve and make data available, was supported.

4.5.3 Open to whom?

An important issue discussed related to whom data would be open to. This central question has a major bearing on how datasets need to be prepared, who decides and the governance of this process. There were two elements that characterised public views on this.

The first was practical: overall it was felt that the vast majority of interest in datasets would be from other researchers. As noted earlier, participants could only see limited demand from the general public to explore raw data – and overall couldn't see the point in placing data out there for the sake of it, particularly if there were significant cost implications. Even if having a medical condition to which a particular dataset related, participants would be more interested in hearing about the results of experiments than having access to the data themselves.

> "I wasn't saying don't put it up there, what I was saying it does not need to go up there for everybody, it can be put up selectively for scientists and people to access, that we the general public who can't interpret the data as such, and don't really necessarily need to do anything with it, then it needn't be up for us, that's what I. I'm not saying it shouldn't be

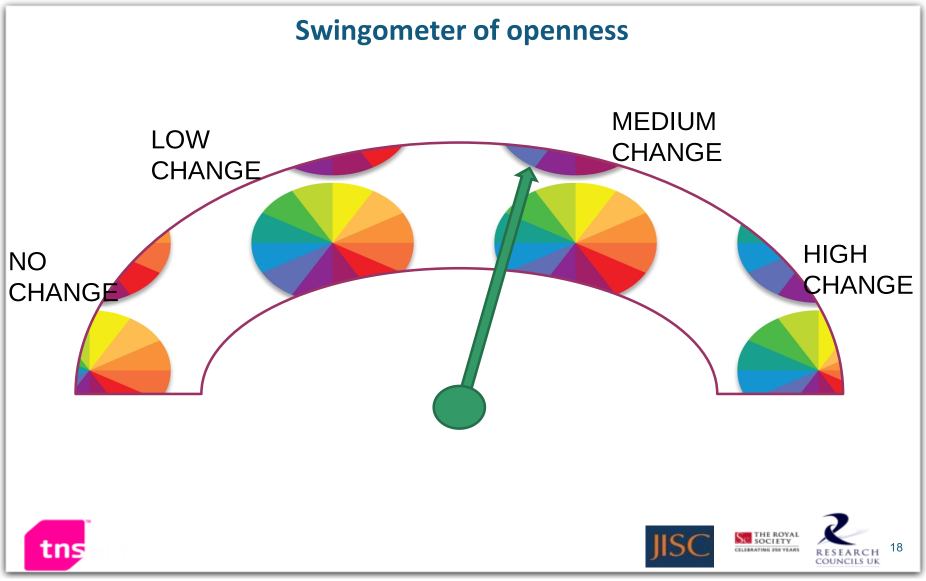
> up there, I'm just saying maybe it doesn't need to be up there for everybody."
> Swindon, Wave 2, Group 1, Female.
>
> "When they talk about openness, I mean, a lot of the data, I won't understand. But I know if it's accessible to the public, it's accessible to other scientists who can validate it. That's where I'm coming from, about openness. Not Joe Bloggs on the internet saying I can look for better research. I'm not going to understand it. But… if it's open, then it's available to other scientists to look at."
> Oldham, Wave 2, Group 2, Male.

The second was in principle: despite not being able to envisage a reason personally, participants felt (particularly when considering case studies) that anyone should be able to access the raw data if they wanted. The governance and data management plans for individual projects should therefore enable as much access as possible, without entailing excessive costs. The balance of this needs to be decided on a case by case basis, and was part of the role of those governing data at Research Councils, universities and other research establishments. Participants also highlighted the idea of trialling open data in certain areas of high public interest before fully rolling it out across all academic research areas:

> "How much it's going to cost and also the relevance to us having open data. I'd like us to have it but I want to be able to see the changes it makes to us before I'd say change it all."
> Oldham, Wave 2, Group 1, Female

Finally, during this session, participants were asked to complete an exercise where they highlighted whether current data practices should change, using a "swingometer" – see Figure 1.

Figure 1: Swingometer of openness



Swingometer of openness

Whilst the absolute position of the arrow was contingent on how open participants currently understood researchers to be, the overall sense was that researchers do need to open up publically funded data sets, but this should be done in a gradual way that did not disrupt the research process. Opinions ranged from low levels of change through to medium/high change. Some indicated higher levels of change where research was considered to be in the public interest.

4.6 Moral and ethical issues around open data

Case studies were used to explore ethical and moral dilemmas.
The public was viewed as specifically having the right to:

- informed consent for personal data use;
- overview and scrutiny of governance arrangements around personal data;
- research data funded by the taxpayer reused for collective gain;
- research data in the public interest disclosed

Researchers were viewed as having the right to:

- exploit data for publications
- reuse public data

They had responsibilities to:

- protect individuals' data and confidentiality
- keep data in a format to enable reuse
- develop data management plans

Funders had rights to:

- expect data management plans to be developed and actioned by researchers

They had responsibilities to:

- develop appropriate governance mechanisms to scrutinise public interest;
- develop contracts with private sector that seek appropriate balance in data openness;
- Enforce and publish data management policies and practices

To explore ethical and moral issues in the workshops, a range of case studies was used that highlighted different dilemmas around data reuse and disclosure. Based on a series of real life examples, and brought to life though the use of actors, the case studies looked at tensions between public and commercial interests, whether data itself constituted intellectual property , the impact of mixed funding on openness, the reuse of personal data, social and commercial benefits, and vexatious requests.

Overall, ethical and moral issues were articulated in terms of rights and responsibilities for different stakeholders. They are summarised in Table 1.

Table 1: Rights and responsibilities around open data.[12]

	Rights	Responsibilities
Public	• Informed consent for personal data use • Overview and scrutiny of governance arrangements around personal data; • research data funded by the taxpayer reused for collective gain; • research data in the public interest disclosed	
Researchers	• Exploit data for publications • Reuse public data • not to publish their data for commercial reasons unless in the public interest [Private sector only]	• Protect individuals' data and confidentiality • Keep data in a format to enable reuse • Develop data management plans • Publish data, within a specified time period [Public sector only]
Funders	• Expect data management plans to be developed and actioned by researchers	• Develop appropriate governance mechanisms to scrutinise public interest; • Develop contracts with private sector that seek appropriate balance in data openness; • Enforce and publish data management policies and practices

Public funded researchers should have an active presumption of publication, within a specified time period.

4.6.1 The public

There were two broad categories of ethical concerns for the public:

- individual rights, in particular privacy, confidentiality and consent around personal data;
- collective rights, which related to ensuring research data in the public interest was disclosed and used for public benefit. It also related to the misuse of data.

---

[12] This framework is based on the *Ethical Matrix* - developed by Professor Ben Mepham, Director of the Centre for Applied Bioethics at the University of Nottingham.

With regard to individual rights and how best to protect personal data, privacy was very important. In principle it was believed that personal data should be anonymous where possible. In this context the majority of people were open to data reuse; and where it is not possible, reuse should be subject to gaining consent.

> "Well, if you're anonymous and you're only XYZ123 it doesn't matter, does it? I think it's only if you're giving your names and other information and personal data."
> Oldham, Wave 2, Group 1, Female

In practice gaining consent was seen to be difficult for the following reasons:

- The nature of consent may be conditional and it is hard to imagine all possible reuses of data, some of which may conflict with people's values.
- Combining datasets may both mean that the ownership, together with rights and responsibilities around reuse, becomes unclear.
- Related to this, combining datasets may impact on privacy issues by revealing information about individuals - for instance combining administrative, health and other research data which could be used to effectively draw conclusions about a single household or individual.
- For certain data sources - such as NHS data, as well as administrative data – there is no 'opt in': it is collected at point of use.
- Standards governing potential multiple reuse was an issue - particularly if one company was to sell information on to another.

With regard to collective rights, the primary ethical issues related to opening data to improve quality of life, in particular by improving health outcomes, as well as things such as environmental improvements. Defined as research in the greater good or public interest, generally (and subject to enabling firms to realise an appropriate return on investment), such datasets should be opened in both public and private sectors. This public interest argument generally trumped all other ethical concerns – this in itself has ethical implications for those overseeing open data practices, as this conversation from Oldham group 1 illustrates:

> M: "It's the greater good. It's the greater good of everybody.- You need to break eggs to make an omelette. So, basically, some people are going to object to it [realising personal data]"
> F: "How do you know what the greater good is if you're not told what your stuff is being used for?"
> F: "The greater good as judged by whom?"

> Oldham, Wave 2, Group 1

Quality of life was often framed in terms of not doing harm, for example through disclosing negative results to ensure businesses were not hushing up problems with efficiency of particular drugs or treatments, rather than openness to promote innovation.

When considering publically funded research in particular, there was a strong moral argument to enable wider use of data given the current austerity measures in public policy.

Finally, there was also an ethical concern that opening up data sets could increase their misuse, as people interpret statistics out of context, and use it for personal or political ends, rather than for the public good.

4.6.2 Researchers

There were a number of ethical issues concerning researchers, expressed both in terms of rights but also in terms of a series of responsibilities.

In terms of researchers rights, participants strongly believed that anyone investing time, effort and expertise into collecting data should be granted a period of time for exclusive access to that data, to help create value for themselves personally (in terms of wealth, kudos and career), their organisation and, where possible, for society. This sense of fair play overrode the immediate requirement to open data – with participants concerned that third party researchers didn't get credit for the hard work of others: described as '*creaming off their results'* and '*trying to get all the glory'*. This period of time was seen to be contextual to the research process and set within data management plans.

> 'But you can't set a time…. after 3 months this happened, after 6 months, because the initial data collection might take a lot longer, it almost needs to be like a specific stage that's identified within the process rather than a time constraint'.
> Swindon, Wave 2, Group 1, Female

Related to this, participants were concerned around the impact of vexatious requests for data on researchers. Though this is covered in data protection law, overall the development and publication of data management plans were felt to help mediate this, and the idea of payment for requests outside of this timetable was suggested to help prevent abuse.

> “If those [data management plans] were in place, it would ease the pressure so if someone came and said, okay, I want everything right now. Well, you know what? I’m not going to spend £60k just to please you. You know? That’s a waste of money if everyone came to me and did that”
> Oldham, Wave 2, Group 1, Male.

In addition to rights around their own data, all researchers were believed to have rights to access public data. In a case study which explored reuse of data by a tobacco company, which had originally been collected to encourage teenage smokers to quit, participants did not believe access to the data by the company should be restricted, providing necessary consents were gained from the research participants. Overall, openness was strongly associated with being open to all – irrespective of whether purposes related to the public good (with the notable exception of data relating to national security).

Beyond these rights, a series of responsibilities was highlighted. These included the need to ensure that data is accurate, and to keep data in a format to enable reuse and develop data management plans. Whilst important, as noted earlier, participants wanted to ensure that researchers did their day job, rather than using significant amounts of time on administration and data curation. It was noted that specialists in this area could help to enable this to happen more efficiently.

However, the most important responsibility was to protect individuals’ data and confidentiality. Though this issue is already covered by much regulation, participants were concerned around the trustworthiness of different organisations to do this effectively.

Finally, there were differences between publically and privately funded researchers. Specifically, privately funded researchers had the right to not publish their data for commercial reasons unless in the public interest. This was seen very much as the fabric of commercial innovation, and participants could not easily see the motivations for businesses getting involved in funding research if data was then opened to competitors.

> “As long as there are people who have a commercial interest and who have taken the risk of putting the money up front, they should get some protection and should get the investment that any sort of profit brings. It’s a commercial decision for them to go into these things; if they take the risk and put their hard-earned money in there, they deserve protection.”

> Swindon, Wave 2, Group 2, Male.

Conversely, public funded researchers should have an active presumption of publication, within a specified time period.

4.6.3 Funders

Funders and other gatekeepers with a commercial or business interest in the research data (for instance a university or company) were seen as pivotal for good governance – not only in terms of policy development but also in terms of creating the culture of openness in which the research was undertaken. In particular, they had a number of significant responsibilities.

Foremost amongst their responsibilities included reviewing whether different areas of research were likely to have strong public interest elements, and then developing appropriate structures to decide on the appropriate release of data.

The issue of co-funded research between the private and public sector was discussed in this context. Whilst, as noted earlier, there was a strong view that commercial organisations had the right to protect their Intellectual Property, overall joint partnerships should seek to balance the range of interests, rather than stacking the cards to benefit commercial organisations. It was deemed very important to set up contracts or model agreements where the rights around data reflected the balance of financial and in-kind contributions of the various parties. There was a perception that in the desire to commercialise research and work with business, universities may be under pressure to give up too many rights, which in turn could affect the public interest – as this conversation from Swindon illustrates:

> M: "It does feel a bit as though we're being cheated, because its research for hire, we're getting a commercial company to tap into the integrity of our universities to do all their research only for then the company to probably make maximum profit from their findings. I mean great for the kudos of the university but I'm not convinced that there's much more out of it than that is there at that stage?"
> M:There is a certain percentage of really, but for them to have a say in it and keep it closed they should have a high percent stake in it, so if they said if it's below a certain amount then they've got no choice."
> F: I think if a commercial enterprise has funded something then they should have the benefits from it, and how do you draw the line between if they put a third in how much do they get out of it, you can't just draw the line and say well they put more in than the government, they should have more out. If they put some in and it's to be commercial for them,

> and they're not going to put the money in unless there's something in it for them anyway as a private firm.”
> M: “So why not do it proportional to what they put in.”
> F: “But how do you proportion the data?”
> Swindon, Wave 2, Group 1.

In this context, it was also noted that researchers often uncovered things they were not expecting to find. In this context, contracts or agreements have ‘what if’ provisions to protect the interests of various parties.

The final responsibility for funders related to publishing data management plans, so there could be scrutiny of the governance around data openness. This could be tied to other information about projects which would be of wider interest – including what the research is about and why it is funded.

Finally, funders should police this process, so it does not become a tick box exercise. In this context, funders had a right to expect that those in receipt of grants would both develop and action data management plans.

> “All of the, everything that’s funded is put on the website so you can see the extract, but you wouldn’t, so the little summary of what the research is about, that’s all publicly available, not when the data would be available it's not disclosed at the moment, and it's not always decided upon.”
> Swindon, Wave 2, Group 1, Male

4.7 How data could be governed

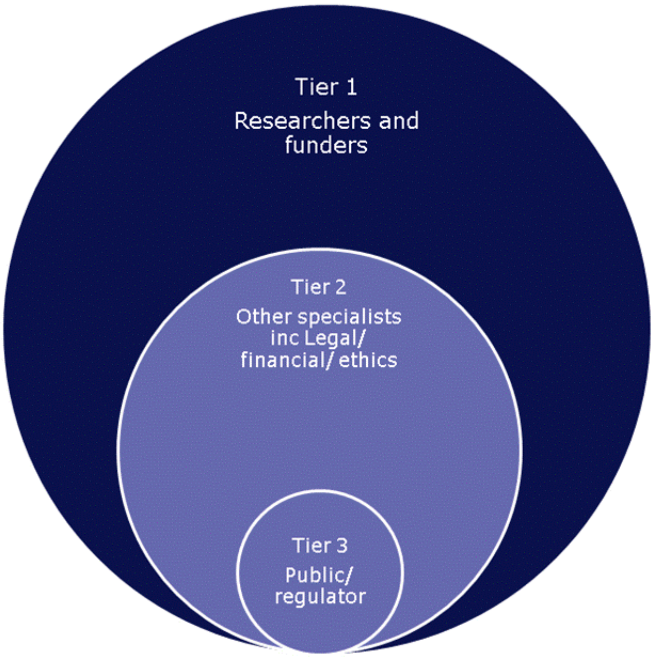
Participants were asked about the arrangements that would need to be put in place to effectively manage open data. Governance of data was strongly related to the perceived level of public interest in the research. Three levels were distinguished:

  - Tier 1: Data with no public interest implications;
  - Tier 2: Data with public interest implications; and
  - Tier 3: Data with public interest implications that includes personal information (see Section 5.1 for further information)

- Differing governance structures were discussed according to the level of public interest.
- Overall, there was a strong sense that any governance arrangements should not be too cumbersome and focus on enabling researchers to do research rather than administrate.
- There were no clear views around the extent to which current open data governance arrangements in the private sector were adequate

As noted earlier, the governance of data related strongly to the perceived public interest in the research. Broadly three levels were distinguished: data not in the public interest (tier 1); data with public interest implications (tier 2); and data with public interest implications that includes personal information (tier 3) - see Figure 3.

Figure 3: Possible governance structures for data



Tier 1 was associated with data with no public interest implications– and was generally felt to apply to most research. Data governance would be overseen by funders and other researchers and funders, who have appropriate technical knowledge and understanding

Tier 2 was associated with data with public interest implications, such as that affecting human health or where there are significant implications for the environment. Here data governance would involve specialists including those from ethics, legal, financial, faith backgrounds and so on. But it would generally not include the public.

Tier 3 was associated with public interest research with implications for personal data. Governance could involve the lay public as well as specialists. Regulators should also govern this area.

Overall, there was a strong sense that any governance arrangements should not be too cumbersome and focus on enabling researchers to do research rather than administrate. The need to mandate the disclosure of academic data that was felt to be of marginal relevance to the public interest was also questioned by respondents as it was likely to entail costs that could be better spent on research.

These views need to be seen in the context of very low awareness of the research councils and their role. When participants learnt that the research councils administer grants to fund public research and that they could help provide both an overview and scrutiny of data, this shaped participants views. They considered that the Research Councils met the requirement (in lower public interest instances) of the independent group participants wanted to see governing data release.

Finally, there were no clear views around the extent to which current open data governance arrangements in the private sector were adequate, mainly because practice was viewed as very mixed and there was not an institutional focus for research funding – such as the Research Councils.

As an artefact of the presentations, participants were most familiar with the governance around pharmaceutical data, which they generally viewed as 'good' as it focused on clear data release plans. They felt partly reassured due to the presence of regulators in this area. However, this was tied to a strong belief that commercial interests and profit would shape data release policies, irrespective of public interest arguments. In this context, the focus became on ensuring governance would not lead to withholding data that could cause harm – particularly negative results concerning healthcare.

5. Conclusions

5.1 Defining public interest in research data

There were two broad views on what constitutes the public interest around data reuse that emerged from the dialogue, which have distinct and potentially conflicting implications for data governance.

The first was an instrumental definition and directly relates to whether data has public interest implications. Here data was believed to be in the public interest if it provides a direct utility or benefit. This was generally defined in terms of a health or environmental outcome. The overriding goal when opening data was to promote these outcomes. The governance implications in this context are to open up datasets as soon as possible.

The second relates to a public interest definition that focused on doing no harm. Akin to Hippocratic principles, here data release needs to avoid poor outcomes – defined as things like confusion, misinterpretation or disempowerment. Governance in this context becomes more paternalistic, with data release contingent on the quality of the evidence, timing, whether research has been replicated and so on.

Both characterisations were highlighted during the dialogue, and aspects of both need to be reflected in governance arrangements. Certainly, participants were keen that inaccurate data should not be placed in the public domain. The scope and limitations of the data also need to be highlighted to avoid misleading interpretations.

On balance, however, there was a presumption of active publication of publically funded or public interest data. As such the first definition should be the driving principle behind data openness, whilst mindful of the negative consequences of publication.

5.2 Confidentiality, consent and the greater good

Opening data in the public interest can involve direct trade offs for the public – in terms of individual rights around consent and confidentiality and the greater good, for instance through the secondary use of medical data. Whilst certain participants were keen to gain consent each time personal data was reused, this may be very difficult to achieve in practice given complexity of data linkage, ownership and wider governance. One possible approach is to consider general permission for ongoing consent around anonymous data, and respecting confidentiality and autonomy where participants are identifiable – even if consent is not always possible.

Related to this, greater good arguments – particularly around public health outcomes – were seen to override individual rights in many contexts. Specifically, the purposes that research is seen to serve play an important role in shaping public attitudes towards the use of personal data.[13]

Given the potential to exploit this, dealing with it requires integrity from those overseeing the governance of data not to overstate claims around research, particularly on controversial or contested areas, and to appropriately consider rights of participants as well as potential beneficiaries.

5.3 Who gets to define the public interest

There are a number of implications emerging from the dialogue about who gets to define public interest. In part this will be defined by disciplinary area – with certain avenues of research (most obviously medical) likely to be by definition in the public interest.

Beyond this it is worthwhile thinking about how the current administration of research grants can be used when thinking about data governance. Currently, grant applications routinely consider impact, as well as research ethics and the wider implications of the individual project. In turn, these are assessed by funding committees and, where significant issues emerge, other governance committees (such as ethics or science and society committees).

These existing processes could be used to help categorise research into areas with higher public interest, which in turn may need distinct practices and structures around data governance – namely accounting for public principles around openness and the participation of lay people.

5.4 Building public principles around openness directly into the policies of research funders

There was a strong resonance between views of the public and views on openness emerging from the research councils, with findings in particularly complementing RCUK Common Principles on Data Policy.[14] A focus could now be on improving, codifying and reviewing practices to ensure these aspirations are met.

---

[13] This resonates with emerging findings from the Administrative Data Taskforce Public Engagement Expert Group

[14] See appendix 2.

It should be noted there was perceived to be a significant gap between these principles around openness and current practices – and as such, conclusions from this dialogue should not be read as ‘business as usual’ for researchers. Rather participants were keen to see the piloting and testing of data openness principles in areas of high public interest, before being rolled out further – with particular analysis of the costs and benefits around data curation.

In this context, there was concern as to whether researchers have the time or skill set to adequately prepare datasets for reuse. One option could be to develop specific roles and funding streams for data curation, trialled in the first instance around the most valuable data sets

One particular issue is co-funded research and data rights between the Research Councils, University and business. Particular attention should be given to the relationship between data and Intellectual Property, and how this is enshrined in contracts and model agreements (particularly around exclusive exploitation) in a way that meets the spirit of the principles.

5.5 Potential open data governance structures for the research councils

As highlighted in the dialogue, there is an opportunity to experiment with new structures around data governance. In research areas with high public interest (such as relating to impacts on human health and the environment, or national security; or where personal data are involved) lay or wider professional involvement should be considered. Lay involvement may be particularly important in instances where it is decided not to publish data due to public harm arguments; or where individual and collective rights are in conflict.

Research data with lower public interest (defined as not relating to human health, security and environmental impacts) were believed to be adequately overviewed by researchers and the research councils. In this context, the costs of data curation need to be considered relative to the benefits – as there were significant concerns around wasting public money curating data with minimal reuse value.

A relatively quick win for the research councils is to actively publish data management plans. There was a view that if these are to be effective, they need to be policed - with sanctions potentially imposed on researchers for non-compliance.

5.6 Open data and the wider governance of research

Finally, it should be noted that open data did not capture participants' imagination and overall the issue was considered of greater relevance for researchers than public *per se*. Given also the complex relationship between openness and trustworthiness, addressing open data alone is unlikely to have a major impact on governance concerns around research - which generally centre on things such as the motivations of scientists, who is funding the research, the speed of research relative to regulatory capacity, the handling of uncertainty and so on.[15] In the dialogue a number of these issues were touched on as areas where greater openness could help promote scrutiny and accountability of publically funded research.

Overall, open data, though not unimportant, is unlikely to be a highly significant public issue around the governance of research. If the goal is to help promote trustworthiness in this regard, using the debate around open data to promote reflection by institutions on other areas for open governance would be worthwhile.

---

[15] Chilvers, J. & Macnaghten, P. (2011). The Future of Science Governance: A review of public concerns, governance and institutional response.

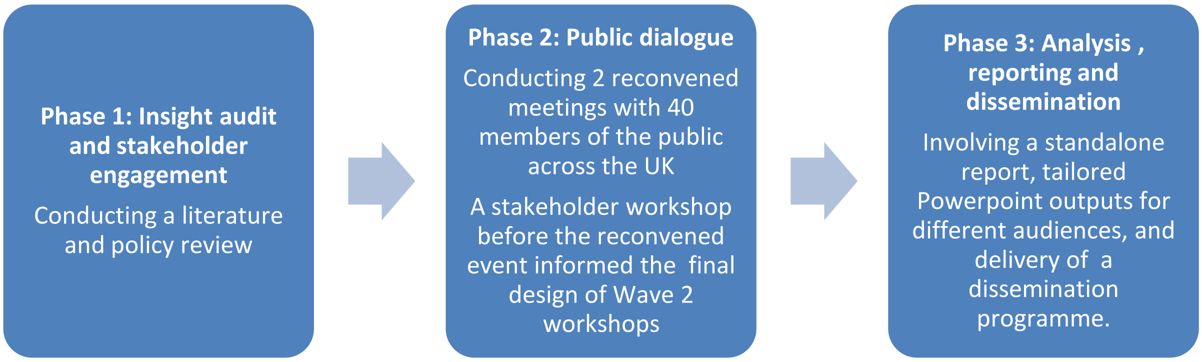
Appendix 1: Method

In this section, we provide details of the method

This is an important study for the research councils and a wide range of influential science, policy and business stakeholders who will be interested in the outcomes of the research. The public dialogue was designed to:

- Provide insight on the business issues that the dialogue will support, at the research councils and JISC
- Build on prior work in the area and account for the wider policy framework
- Engage people meaningfully around this complex area, enable the public to frame issues, be responsive to their concerns, and test out any principles emerging across a range of research contexts

To address this, we developed the following 3 stage approach:



Phase 1: Insight audit and stakeholder engagement

Conducting a literature and policy review

Phase 2: Public dialogue

Conducting 2 reconvened meetings with 40 members of the public across the UK

A stakeholder workshop before the reconvened event informed the final design of Wave 2 workshops

Phase 3: Analysis , reporting and dissemination

Involving a standalone report, tailored Powerpoint outputs for different audiences, and delivery of a dissemination programme.

Phase 1: Insight audit

This phase comprised two components:

- a set up meeting; and
- a policy and information review.

Set-up meeting: this meeting between the full TNS-BMRB team and key people from RCUK, JISC, Sciencewise-ERC and Royal Society, was designed to achieve three goals:

1. To fully understand the decision context for the dialogue, how the findings would be used, and what relevant information already existed for the study to build on, including sensitivities around the project, and what success would look like.
2. To review the methodology in light of this context and agree any substantive changes to the design or scope of the study.
3. To agree a process for mapping stakeholders and to consider who else needs to be involved to make the project a success, together with the hooks to encourage them to take part in the study.

Policy and information review: Following the set up meeting, a rapid review of existing evidence and information around the principles and practices governing open data was produced. Information was drawn from three main areas:

1. Current policy documents at RCUK, JISC, HM Government and elsewhere which govern current practices around data openness and transparency.
2. Findings from previous/current studies and working groups on open data practices. This included a review of evidence submitted to the SAPE group (where not confidential), other RS studies (such as *Science in the public interest*); as well as the principles developed by the Public Sector Transparency Board, and the Technology Strategy Board views on the implementation plan of the Open Data Institute.
3. Previous research on public views in this area. These include reviewing public dialogues on stem cells, synthetic biology and, most importantly, the Science and Trust study which looked in depth at organisational practices around governance, including openness and transparency.
4. To collect their wider views and perspectives on the public dialogue.

Phase 2: Public dialogue workshops and stakeholder engagement
This phase comprised of three components:

- Wave 1 of the public dialogue;
- Stakeholder mapping and scoping workshop
- Wave 2 of the public dialogue

Wave 1 of the public dialogue: The public dialogue was designed to achieve 4 goals:

1. Engage a cross section of the public and make the issues accessible to people
2. Enable participants to help frame the issues for discussion and ensure the process is responsive to their concerns
3. Enable people to develop a series of principles to govern data use and management
4. Explore how these principles play out across a range of different institutional settings and interests

To do this, our approach involved using evening workshops of 2.5 hours and a reconvened approach for this study, rather than one day long workshop. We do this because, from our experience: the subject area is complex and people need time to digest the information. Two workshops, one in Swindon and the other in Manchester took place at both Wave 1 and Wave 2.

The objectives of the Wave 1 workshops were to:

- Explore top of mind views around openness and how it relates to research
- Introduce and inform people about the data openness
- Enable people to think about principles governing data openness
- Develop a list of key principles to be explored in the next workshop

Wave 1 involved two workshops, one in Swindon and the second in Manchester in late February 2012. The discussion guide for Wave 1 workshops is presented at Appendix 2.

The areas covered and the design and stimulus used is summarised below.

Wave 1 – Open data principles

Issue for debate	*Design and stimulus used*
Understand peoples' views about the benefits and downsides of being open.	Get people to think about the positive and negative associations linked to "being open in your own life"
Understand their views about data, and open data. Ask them to think about the benefits and any drawbacks of open data.	Get people to think about an area of science where knowing more information/data would have been helpful (e.g. recent issue on breast implants; MMR; smoking etc). Consider wider benefits of sharing information, and any concerns.
Introduce people to scientists discussing open data and data management	A small ethnographic film/video diary made by researchers/scientists with them explaining what happens to the data they collect over the course of research project.
Illustrate how research is done and what happens to information collected in studies; together with current ethical, legal and commercial constraints; and current quality assurance and peer review mechanisms. **Get people to consider their initial thoughts around principles to govern data**	Questions including: What do people want from data; is there a 'right to data access'; how do you promote trust in data use; what is appropriate to ask of researchers
Get people to reflect on principles	Development of principles to govern data and initial feedback
Consider information or perspectives people would like to know more about	Key issues board and further information requirements for next wave

Between Wave 1 and 2 workshops people were asked to research the issues and speak to friends and family about open data. This enabled us to reach a wider sample of people indirectly, and provide insight into impact of external information on views.

Between Wave 1 and 2 a stakeholder mapping and engagement exercise was undertaken to understand stakeholder needs from the dialogue; their views on the SAPE findings; and gain feedback on the materials and approach for the dialogue.

Stakeholder mapping and scoping workshop: In conjunction with the Steering Committee we identified key stakeholders, and designed a scoping workshop with over a dozen organisations. The workshop was held on 23rd February 2012 at the offices of TNS-BMRB, in London.

The objectives of the workshop were:

- To inform stakeholders about the early findings from the discussions with the public on open data,
- To get their input into draft content and materials designed to use in the second wave with the public, and

Wave 2 of the public dialogue: The objectives of Wave 2 workshops were:

- To answer any questions about open data arising since Workshop 1
- Understand and discuss how open data works in practice in different settings (private and publicly-funded research)
- Explore different perspectives on open data using case studies
- Review the principles governing open data
- Consider who need to be involved in thinking about the public interest in open data policies

Wave 2 involved the same approach: two workshops, one in Swindon and the second in Manchester in early March 2012. The discussion guide for Wave 2 workshops is presented at Appendix 3.

The areas covered and the design and stimulus used in the second wave is summarised below.

Wave 2 – Open data practices

Issue for debate	*Design and stimulus ideas*
Recap on issues emerging from Workshop 1	Reflections from the pre-task and review of the principles
Input from experts on publically and privately funded research on current and future open data practices	Two presentations from experts from the Stakeholders workshop who discuss current and future practice on open data in their organisations.
Consideration of real life scenarios in which open data practices can be further considered.	Use of 4 case studies which actors play the different characters. The scenarios on: publicly funded research; private funded with public implications; research which may impact on individuals' privacy; Intellectual Property and commercial exploitation of data; data reuse etc
Review and conclusions, principles and implications for models of openness	Using a swingometer to assess how they feel about changes in open data practices. Discussion of principles in terms of importance; review of openness models and impact on trust
Governance arrangements and responsibilities of different individuals and bodies to uphold these standards.	Final discussion and wrap-up exploring views on governance arrangements, how these would work, and who would be involved.

1.1.2 Sample and recruitment for the workshops

40 people were recruited to participate in the study in 2 locations: Manchester and Swindon – the latter at Research Council's offices.

A total of 40 participants were recruited. In each workshop a mixture of participants in terms of gender, socio-economic group; age and ethnicity were recruited – reflecting the profile of the local areas of recruitment. Participants were recruited using free-find methods (off the street) and were paid an incentive of £90 for attendance. Incentives were only be paid at the end of the 2nd workshop to promote attendance.

1.1.3 Analysis, reporting and dissemination

The workshop discussions were digitally recorded and transcribed before an analysis of the data undertaken. The analysis consisted of a combination of content and framework analysis using the following processes:

- creative discussion and brainstorming among the research team, with reference to transcripts and recall of fieldwork, to identify key themes
- analysis of notes made on the research materials (e.g. principles cards, swingometer and so on), to identify which issues are perceived to be most important by the public and how this relates to data management policies and procedures.

Final outputs included:

- Presentation of the findings
- Draft and Final Report of the findings

At the time of writing, dissemination activities were being planned which include feeding back to the Administrative Data Taskforce.

Appendix 2: Stakeholder Workshop Agenda

RCUK Open Data dialogue

Stakeholder Workshop 23rd February, 10.00-12.00pm

TNS-BMRB, 6 More London Place, London, SE1 2QY

Agenda

Stakeholder workshop objectives

- To inform you about the early findings from the discussions with the public on open data
- To get your input into the content and materials we have designed to use in the second wave with the public
- To collect your views and perspectives that will be important for the dialogue to consider

10.00	**Introduction to the workshop** - Welcome and participant introductions - Background to open data study - Overall objectives and outputs of the study - Objectives of the stakeholder workshop - Workshop agenda	**TNS-BMRB**
10.10	**Presentation of the Findings from the First Wave of the Public Dialogue** - Public framings of openness - Public framings and questions around open data - Benefits and drawbacks of open data - The Principles of open data	**TNS-BMRB**

10.20	**Plenary Discussion – Stakeholders' views on open data** - Discussion on initial views on public's views, any surprises? Anything unclear? Anything that needs verifying/checking? - How does this compare with stakeholders' views? - What specific areas do stakeholders consider need to be covered in wave 2? Moderator produces a flipchart to include ideas to be covered.	**ALL** **PLENARY**
10.45	**Overview of public dialogue WAVE 2** - Presentation of overall structure of public dialogue session - Presentation of planned scenarios and introduction of task to discuss the case studies	**TNS-BMRB**
10.50	**Break**	**ALL**
10.55	**Group Discussion of Discussion guide and scenarios** - Participants will be divided into 4 groups each with a TNS-BMRB moderator - Each group will be given 30 minutes to discuss the guide and scenarios, thoughts about content, any suggestions/improvements? - Moderator to capture views on flipchart and display on walls	**ALL** **MODERATED** **GROUPS**
11.35	**Plenary feedback on scenario discussion** - Key point presentation back to full group - Any further comments	**ALL** **PLENARY**
11.45	**Engaging with participants around their and other questions** - Importance of stakeholder attendance at public dialogue sessions - Identifying whether stakeholders could attend public dialogue sessions	**TNS-BMRB**
11.55	**Next steps** - Next steps following wave 2 of public dialogue - Thank and close	**TNS-BMRB**

Appendix 3: Open Data Workshop 1 Discussion Guide

RCUK Open Data dialogue
Workshop 1 Topic guide

Research objectives

- To provide public insight and feedback on future data openness, data reuse and data management policy options
- To explore wider ethical and moral issues related to open data and data reuse, for instance the relationship between openness and how the research was funded; and how we balance the interest of different groups about data

This workshop aims to:

- To explore top of mind views around openness and how it relates to research
- Introduce and inform people about the data openness
- Enable people to think about principles governing data openness
- Develop a list of key principles to be explored in the next workshop

		mins
1.1 Welcome and introduction • Introduce TNS-BMRB – independent research company • Introduce research – a study exploring public views on open data. Specifically: under what circumstances is it good to be open about the data created through research? What is the value of this? How do we balance the interests of different groups around the data? • Project undertaken for the research councils – who fund medical, scientific and social research, as well as research in the Arts and Humanities. For instance this can include research on things as diverse as climate change, cancer, art and literature, food, engineering and so on.. • Will also inform a project on this issue by the Royal society – the national academy of science for the UK.	RC rep to say a few words	

○ We will be meeting over 2 workshops ○ First will consider what principles should underpin researchers approach to open data ○ Second will explore how these principles may be applied in practice – using real life case studies ○ Not expecting you to know lots of things about research, data or the research councils. There is no right or wrong answer • Explain format of evening - a mixture of small group and plenary discussion. • Length of discussion approx. 2.5 hours • Housekeeping: mobile phones, toilets; fire etc		
Session 2: Group introductions and framing openness	**Small groups**	**15 mins**
2.1 Group introductions and icebreaker • Recap on goal • Ground rules in group • Gain permission to record group ○ Confidentiality – their views will be used, but not identifiable ○ Recordings only available to the research team • Participants introduce themselves to the group ○ Name ○ Where they are from Before we think about being open around research data, we want to begin by thinking about being open in your own life. • Thinking about someone you know who is very open ○ What are the benefits of the person being very open? ○ Can you think of examples when there are problems in the person being very open • What things affect how open you are? ○ Probe other people ○ Probe - how you are communicating (online;	Note benefits of openness on green cards Note problems of openness on red cards	

face-to-face and so on) • Based on what we have just discussed, we now want to develop some principles on openness. To help us do this: ○ When would you say it is good to be open? ○ When would you say it's ok to be closed?	Note principles on blue cards and place on walls	
Session 3: Initial thoughts on science, research and openness	**Small groups**	**15 mins**
3.1 Initial views on openness and research • In general terms, how open do you think scientists and other researchers are about their research? ○ Probe: What sorts of things have they been open/closed about • What do people think about when we say scientific or research data? [Clarify DEFINTION – HANDOUT Data is the results of measuring something: such as the statistics from an experiment. For data to become information, it must be interpreted and take on a meaning. For example, the number of people in Swindon/Manchester with asthma is generally considered as "data", understanding whether this differs from London and the reasons for this is "information" and "knowledge". The knowledge drawn from data can be uncertain. In the example, uncertainty is increased if we only ask a sample of people in Swindon/Manchester about asthma (rather than everyone). What causes the differences in asthma rates can also be hard to demonstrate, particularly if there are many potential factors. For instance, it may relate to a combination of things such as levels of pollution; whether the patient has had other respiratory illnesses; whether there is a family history of asthma; whether	HANDOUT 1 DATA DEFINITION [Use and adapt cards from previous session]	

someone smokes and so on. It can be hard to prove cause and effect • Can anyone think about an example where there has been an issue around how open scientists and researchers are about research data? ○ (Open – if not probe: a health issue such as breast implants, MMR or smoking; an environmental issue such as climate data). • Thinking about [select one example] ○ What would have been the wider benefits of sharing information? For whom ○ Are there any problems around being open? For whom ○ How do you think some if the principles we have just talked about apply to this issue? [reword principles as appropriate others] • Are there other principles that would be useful when thinking about researchers being open in this context • Thinking about [select the example] does anyone know about the type of data that may be collected in a research project of this nature? •		
3. What do we mean by open data		**15 mins**
Video plenary session Film by researchers concerning the data they collect, and who has access to this, and how data is controlled Also to cover their views on how increased openness may : • help them • hinder them Q&A	Video of 2-3 researchers	
Tea Break		**10 mins**

4. The life of data		**15 mins**
• Initial thoughts and reflections on the video • View on positive and negative impacts of openness. SHOW: possible data and control stages in a research process are [Handout the ‘data journey’ map] ○ Set up ○ [if needed] ethic approval ○ Data collection and analysis [primary] ○ Analysis of other data [secondary] ○ Publishing ○ Data reuse ○ [Note it is a process with other people building on data collected] • Does anyone have a further questions or points of clarification?	Data Journey Map [Handout 2]	
5. What do people want from open data		**30 mins**
• How open do you expect researchers to be about the data they collect? ○ What is appropriate to ask of researchers collecting data? • • Should data be open for anyone to scrutinise? ○ What is the benefit/shortcomings of this • • Does it make a difference if the research is? ○ public funded ○ private funded ○ has public implications ○ information relating to privacy of individuals • • What would be the impact of greater openness? ○ Unprompted and probe trust • • Here are some things people have said about open data [show speech bubbles 1 per time - 5 max] ○ How do you feel about this viewpoint? ○ Does it change your views about what you want from open data?	Speech bubbles from different groups [Handout 3]	

•		
6. Principles around open data		**30 mins**
We have developed [10-12] principles around open data, based on the things you mentioned earlier [blue cards] and things that other science organisations have said about this [yellow cards – pre-prepared] • Is there anything important we have missed? Now, we want you to rank them high, medium or low, in terms of thinking about how to think about open data in projects. For each ranking probe, • Why do you say that? **Thinking about the most important principles [category ranked high]:** • Should the principles apply equally to private and public funded research? • Are there any principles that are important at different stages of the research process? ○ [if time also probe those ranked low or medium]	Hand out of principles cards [Handout 4] Rank (high medium; low in terms of importance	
7. Information needs		**10 mins**
• What further questions do you have about open data? • Is there any further information would you like to help answer these? • Are there any particular groups you would like to hear from	Key issue board	
8. Thanks and pre-task for next time		**5 mins**
Next time we will be exploring some of how some of these principles may play out in practice, using case studies about research Between now and the next time we would like you to: • speak to friends and family about open data • consider further issues where data openness may or may not be helpful [e.g. in relation to a news item you have seen] • come back with 1 key question for us next time We would also encourage you to do your own research between waves 1 and 2 to help you do this.		

Next meeting is on [6 March Swindon/ 8 March Manchester] @ same venue. Check date is in people's diary. Thank and close		

Appendix 4: Open Data Workshop 2 Discussion Guide

RCUK Open Data dialogue

Workshop 2 Topic guide

Research objectives

- To provide public insight and feedback on future data openness, data reuse and data management policy options
- To explore wider ethical and moral issues related to open data and data reuse, for instance the relationship between openness and how the research was funded; and how we balance the interest of different groups about data

This workshop aims to:

- To answer any questions about open data arising since Workshop 1
- Understand and discuss how open data works in practice in different settings (private and publicly-funded research)
- Explore different perspectives on open data using case studies
- Review the principles governing open data
- Consider who need to be involved in thinking about the public interest in open data policies

	Stimulus/ tasks	**Approx timing**
Session 1: Welcome	**Plenary**	**5 mins**
1.1 Welcome and introduction • Welcome and thank participants for attending the second workshop • Review study objectives of the research study: exploring public views on open data. • Project undertaken for the Research Council UK – who fund medical, scientific and social research, as well as research in the Arts and Humanities. For instance this can include research on things as diverse as climate change, cancer, art and literature, food, engineering and so on. • Also funded by the Royal Society – the national academy of science for the UK and JISC who are responsible for driving	SLIDE 2 RC rep to say a few words reinforcing importance of workshop	

information and digital technology in education. • The first workshop examined the principles of open data. The second and final workshop will cover how open data could work in practice. Specifically we will explore: ○ Any questions people have about open data since the last session ○ The pros and cons of open data ○ Listen to the views of those involved in research in the UK ○ Experience some real life examples of open data (using actors) ○ Consider the public interest in relation to open data • Final thoughts and reflections about next steps in dealing with open data • Explain format of evening - a mixture of small group and plenary discussions, Q&A sessions and a role play played by actors. • Length of discussion approx. 2.5 hours • Housekeeping: mobile phones, toilets; fire etc	OPENING SLIDE 3	
Session 2: Intro and Pre-task: questions arising	**Plenary**	**10 mins**
2.1 Recap on ground rules • Ground rules in group • Gain permission to record group ○ Confidentiality – their views will be used, but not identifiable ○ Recordings only available to the research team ○ Reintroduce one another **2.3 Recap on what was discussed last time**	SLIDES 4-8	

First lets recap quickly on what we mean by data, and the different types of data that are included in the scope of the research **SLIDE 4 and 5 – What we mean by data** **SLIDE 6 – Examples of data** Since we last met we held a workshop with senior people involved in considering open data and future plans. We presented what we had discussed in the public workshops and asked them if there was anything that they thought was missing or should be highlighted. One area was the distinction between confidentiality and consent. **SLIDE 7 – Confidentiality and consent** Also, by way of recap, you will remember we asked you about any questions you had about open data – and here is a summary of the issues that were raised: **SLIDE 8 – The questions you asked last time** **2.2 Pre-task** Do you have any other questions you want to add to this list? At the end of the last workshop we asked you to: • Speak to friends and family about open data. How did that go? ○ Probe: What was discussed / what they said to others about the group discussion • Consider further issues where data openness would be helpful [e.g. in relation to a news item you have seen]. Any comments? • Come up with 1 key question. What is your key question? • How do you feel about the idea of open data now? Are your views still the same as they were last time we met, or has anything changed? Have you had any other thoughts or ideas?		
Session 3: Current and future practices around open data	Plenary	**15 mins**
• Following on from your questions we have invited those involved in research to come and speak about their perspectives on open data. • We have two people involved in private and publically funded research to tell us about their views on current and future	Discussion with 2 or 3 researchers/ decision makers; 1 public	

practices around open data • Introduce the researcher / decision maker by name and ask them to say a little about themselves (job role, area they work in, private or public-funded research, interest in open data) • Two research/ decision maker present a five minute perspective on ○ Current cultures and practices around open data from public and private source (and joint funding) ○ Views on the benefits and disadvantages of increased openness around data ▪ Who should data be open to ▪ Impacts of different degrees of making data open Q&A for clarification	funding; 1 private	
Session 4: The pros and cons of open data	Small groups	**25 mins**
In this session we want to reflect on the presentations in more depth and link it back to the conversations from workshop 1. So to kick off with: • What do you think about the current open data policies and practices? • What works well? • What less well? • Should they be changed in anyway? Thinking about this in relation to the conversations we had about whether trust and open data in workshop 1: • Do you think the current ways of making decisions about open data are trustworthy? • Would being more open and allowing others to scrutinise the data make change your views on this? ○ Probe: trust in the data or in the organisation conducting the research? **We now want to touch upon some of the other advantages and disadvantages around open data** **Benefits/Advantages**		

- What do people feel about the other benefits or advantages the speakers highlighted around increased openness of data [NB probe in relation to talk]

- What would you say was the most important; what the least.

- In addition to the presentations, here are some other things people have said about the advantages of being open [SHOW CARDS: eg McKinsey; NASA and google earth; role in innovation]

Cost/Disadvantages

A3 SHOW CARDS

- What do people feel about the costs or disadvantages the speakers highlighted around increased openness of data [NB probe in relation to talk]
    - eg preparing the data, writing background information about the data, and putting it in a format that is easy to access.

- What would you say was the most important; what the least.

In addition to the presentations, here are some other things people have said about the costs/disadvantages of being open [SHOW CARDS: eg costs 5% of the total research budget; etc]

Do the financial and resource costs of preparing the data to share change your views on what should be made available? In what way?

A3 SHOW CARDS

- Does it affect what is shared?
- Who it is shared with?
- What those requesting the information should contribute?

- What would:
    - be the basic minimum level of service include (i.e. a BRONZE LEVEL)?
    - What would a SILVER LEVEL of service
    - What would be a GOLD level of service

- Before we break for tea, I would like you to think about the

pros and cons and by using a voting game let me know whether you are in favour of: ○ **Maintaining the status quo** and having no change in moving to more open data ○ **A small movement** towards an increase in open data ○ **A large movement** towards an increase in open data	VOTING GAME/ SWINGOMETER	
Tea Break		**5 mins**
Session 5: Review of open data principles	**Plenary**	**10 mins**
Review of emerging principles from session one Do people still feel these are important? Are there any additional principles that need to be added? [unprompted then…] I want to specifically get your views on the confidentiality and consent issues we discussed in the introduction. ○ PROBE: Should principles be about data being anonymised or should it be confidential? EXPLORE DIFFERENCES ○ Could data be used for a different purpose? (E.g. what about respondents agreeing to take part in smoking case study having their data used by Tobacco Company? (Oldham) ○ PROBE: Whether archiving is a missing principle In the next session we are going to have some role plays highlighting different aspects of open data. I would like you to think about these principles and consider whether they work or need to be modified as a result of the information in the role plays.	Hand out principle cards	
Session 6: Two case studies on data reuse	**Plenary**	**45 mins**
5.1 Role Play • In this next session we will use actors to role play two different case studies. All the case studies are loosely based on real events, as follows: ○ Swindon – Tree Rings and NHS Database ○ Oldham – Smoking and Flexiforce • In each case study I will introduce the roles that the actors play.		

They will then set out their position on whether they think the data should be open or not. • Then you can ask them questions which they will answer. **5.2 Case study 1, 2 or 3 and 4** • Role play 1 – Use questions in CASE STUDY stimulus • How did you find the case study? Did you learn anything new? How did it make you feel? • What people felt about the arguments ○ What did you think of both sets of arguments? Which arguments were convincing? Which were not? Who do you think is right? Why? Did you find it easy to come to decision? • How does the case study compare with the principles? ○ Do the HIGHLY ranked principles still stand? Are some more important than others? ○ What other principles need to be added? ○ Should they be re-prioritised? • Repeat for Role play 2 **5.3 The decision** [for tree ring/NHS database/smoking/Flexiforce] • In this example, a decision was made where the University was asked to open their data ○ How do you feel about this? ○ SHOW PRINCIPLES CARD: Does this affect what your view of the principles?	CASE STUDY STIMULUS Principles cards	
Session 7: Who decides what's in the public interest		**20 mins**
• Finally, one of the key issues that came out of the first workshop relates to *who decides whether opening up data is in the public interest*. Given the views of the speakers earlier and discussion in the case studies ○ What do you feel about this ○ What role should the people who fund research have? What about the researchers? ○ Who else needs to be involved? ○ How does this differ between public/private/joint funded research? ○ What other things need to be considered to ensure that open data is managed effectively?		
Session 8: Reviewing what we have decided and next steps	**Plenary**	**10 mins**

- What are the key messages you think that RCUK need to take out of this workshop?
- What next steps would you like to see from those funding research in the
  ○ public sector
  ○ private sector
- Any final issues
	FLIPCHART KEY MESSAGES	

Session 9: Thanks and next steps	**Plenary**	**5 mins**
Thank and highlight what happens now		