


Department for Business, Energy & Industrial Strategy

# Big Data Workshop

Report of a workshop held to explore future public engagement in the field of Big Data

**October 2017**

This report summarises a day's workshop held in Westminster on the 22nd September, which brought together policy makers and researchers working at the 'cutting edge' of big data.

A short briefing paper was used to structure the conversations over the course of the workshop. The briefing paper is included in the appendix.

The workshop addressed the following topics, through a mix of individual, table-based and plenary activities:

**1.** Agree a shared representation of the 'field' of big data research

**2.** Take stock of the public engagement that has been undertaken to date and the key lessons learned

**3.** Horizon scan to identify critical 'hotspots' in research and application over the next 5 years

**4.** Identify a strategic response to the above, agreeing how future public engagement activity might be better targeted and prioritised

This report uses the notes from the various discussions to identify the key conclusions and insights gleaned from the workshop.

We offer some prompt questions within the report where we would particularly value your further reflections and advice. These are highlighted in red boxes

# 1. Mapping the Big Data landscape

Big Data is a broad and somewhat abstract term which describes many different areas of activity.

Delegates were presented with a table which attempted to describe some of the key areas. These were interrogated and working individually, on tables and in plenary, delegates refined the categories and questions that were described in the map.

They suggested additions, identified areas requiring clarification and tested the logic and utility of such a representation of the field.

While it was agreed that having such a map could be useful, it was also agreed that representing it in a linear form was not very helpful: the categories overlap and interact in ways which such a representation fails to capture.

In responding to the feedback, we have generated a new representation in the form of 'big questions' about data, included over the next three pages. In preparing these questions we have:

- Added in details to each category which delegates suggested were missing
- Divided the 'big data categories' heading into three new headings ('data categories', 'provenance' and 'accessibility and ownership')
- Re-ordered the categories
- Presented them as a series of 'orientation' questions to allow people to make sense of the landscape and to position their work / interests within it

## The original map

| Domain | Focal points for investment and activity | | |
|---|---|---|---|
| Big Data **infrastructure**: improving the dynamics of data creation, access, storage and use | • Technical infrastructure<br>• Interoperability and access<br>• Analytics<br>• Capability and capacity | | |
| Big Data **categories:** the types of data being researched and applied | • Archive data<br>• Environment data<br>• Research data<br>• Clinical / health data<br>• Longitudinal survey data | • Admin data<br>• Business data<br>• Open data (public sector)<br>• Social media data | |
| Big Data **research domains:** areas of society in which applications of data are being researched | • Health and social care<br>• Services<br>• Government<br>• Energy | • Environment<br>• Creative economy<br>• Business<br>• Transport | • Cities<br>• Civil society |
| Big Data **applications**: areas where data has transformational potential | • Machine learning (including autonomous vehicles)<br>• The Internet of Things<br>• Artificial and augmented intelligence<br>• Decision making and service design | | |
| Big Data **research themes**: focal points for investigation | • Increasing the accessibility and usability of the data<br>• Trust, identity, privacy and security: the ethical dimensions to data capture and use<br>• Social and cultural applications: how data can transform everyday life, from culture to health<br>• Commercial application: exploiting the potential of data to unlock economic benefit | | |
| Big data **stakeholder groups**: the key types of organisations with a stake in the area | • **Research**<br>  • *Research funders*<br>  • *Learned societies*<br>  • *Research leaders / principal investigators* | • **Policy making**<br>  • *National government departments*<br>  • *Other government / policy agencies*<br>• **Societal stakeholder groups**<br>  • *'Watchdogs'*<br>  • *Societal stakeholders*<br>  • *Professional representation* | |

# The big data landscape

The next three pages show how we have taken the feedback on the table to update the categories, and present them differently: as a series of prompt questions, rather than a table. We suggest that these questions (and prompts) provide a useful framework for describing activity in the area. The right hand boxes are not meant to be exclusive, and would be expected to change over time.

We would welcome your help to develop a more engaging & helpful visual representation, and to refine the prompts & questions.

**Big data usage:**
What are the key activities for which big data is being used?

**Big Data usage**

- Research (to inform understanding)
- Archive (to preserve and make accessible)
- Administration (to inform decisions and practice)
- Commerce (to realise commercial benefit)

**Domain of data use:**
In which areas of public life is the data being gathered / researched / used?

- Health and social care
- Services
- Government
- Energy
- Environment
- Creative economy
- Engineering
- Business
- Transport
- Cities
- Civil society
- Manufacturing
- Financial
- Crisis preparedness and response
- Consultancy
- Employment
- Criminal justice
- Advertising
- Education
- Insurance
- Agriculture
- Defence / security

**Big data applications:**
How and where is big data being applied in the 'real world'?

- Machine learning (including autonomous vehicles and robotics)
- The Internet of Things
- Artificial and augmented intelligence
- Decision making and service design – including role of algorithms; prediction
- Healthcare
- Use of sensing data
- Cyber-physical systems (ultra-connectivity, wearables)
- Virtual and augmented reality
- Visualisation
- Understanding, predicting and influencing behaviour
- Content – e.g. broadcasting, news, social media
- Decisions on what data is presented to the user and how


**Categories of big data:** What types of data are being captured and used?

**Categories of data being used**

- Environment data
- Clinical / health data
- Biometric data
- Transport data
- Business to business data
- Commercial customer data
- Financial data
- Linguistic data
- Official statistics
- Identity / authentication data
- Legal data
- Management information
- Public attitudes to data / use of data

**Big data provenance:** What is the source of the data?

**Data provenance / source**

- Personal / non-personal data
- Primary data
- Derivative data – from analysis of primary data
- Aggregated / pseudonymised data
- Crowd-sourced data
- Messy / heterogeneous data
- Longitudinal survey data
- Social media data
- Transactional data
- Uncertain data
- User-generated data
- 'Shed' data that the public are generally unaware they are creating

**Big data accessibility and ownership:** Who owns the data and how accessible is it?

**Data accessibility and ownership**

- Open data
- Closed/private data
- Passively generated data with low public awareness. Who knows to ask for access to it?

**Big data research:** What are some of the big questions that researchers are trying to address?

- Increasing the accessibility and usability of the data
- Trust, identity, privacy and security: the ethical and legal dimensions to data capture and use
- Social and cultural applications: how data can transform everyday life, from culture to health (positively or negatively)
- Commercial application: exploiting the potential of data to unlock economic benefit
- Systems science
- Public attitudes, awareness and understanding of data
- Understanding the historical and social context for data

**Big Data infrastructure**: what underpinning technical and legal systems and , skills and processes are needed to ensure good use of data?

**Big Data infrastructure**

- Technical infrastructure
- Interoperability and access, metadata
- Ethical frameworks for data collection / use [which will influence the next three]
- Capability and skills
- Governance and regulation
- Education and engagement

**Big data stakeholders:** Which societal groups have a particular stake in the area?

**Research**

- *Research funders*
- *Learned societies*
- *Universities and research institutes*
- *Industrial R&D*

**Policy making**

- *National government departments*
- *Other government / policy agencies*
- *Local authorities / LEPs*
- *Regulators*

**Societal stakeholder groups**

- *'Watchdogs'*
- *Business / industry*
- *Private / public*
- *Citizen scientists*
- *End user groups*
- *Media*
- *Education sector*
- *Campaign groups*
- *NGOs*
- *International*

# 2. Taking stock of public engagement with big data

Delegates were then invited to take stock of the existing and planned public engagement activity in the area. The conversation focused on three topics, and reviewed the available evidence:

**1. Social intelligence**

**Do we have the mechanisms in place to ensure we are capturing and sharing effectively intelligence about public views, attitudes and understandings?**

**2. Engagement activity**

**Are we investing in a judicious mix of high quality engagement activities to maintain open, two-way interaction with the public?**

**3. Engagement capability**

**Do we have the skills and capability to commission, undertake and evaluate effective public engagement, in the policy, practice and research communities?**

There was broad agreement on the three headline points which emerged:

**The pace of technical change** is outstripping our ability to comprehend and consider the social and ethical implications of developments in data. This is a profound challenge and makes a more concerted investment in public engagement an urgent priority. Business will step into the void and set the agenda if government and research don't step up

**Data literacy**: while dialogue is vital, this needs to be built on a secure foundation of understanding. There is an urgent need to enhance our collective efforts to ensure that the public are supported to make sense of and comprehend the nature and potential impact of developments in this area.

**Silo working**: delegates identified a problem with 'silo' working, with a lack of join up and alignment across different areas. There was enthusiasm for more cross-sector, cross-domain working, and for interventions like the draft Big Data landscape map. It was agreed that innovation and economic potential lies in unlocking these silos and focusing on the touchpoints between them.

Focal points for **collective activity** included the following:

- Sharing learning and intelligence more competently across the different 'siloes' of activity
- Better use and understanding of evaluation
- Better coordinated approach across 'interested parties' to support 'big dialogues' happening – but also flexibility to allow more responsive and small scale dialogues

# Other points included

**Challenges in progressing public engagement with big data**

- We lack collective leadership and vision. It would be great to have more people standing up and championing the power of data for 'good', and to convene conversations that focus on tensions and trade offs.
- Currently, there is lots of low level engagement, we don't know what is happening and will need to develop capacity / capability if we think this is something we should know more about.
- It is critical that activity in this area is principled: delegates stressed the need for explicit ethical principles to underpin interventions, focused on integrity and social justice
- Big data is a tool – is it too big/broad a frame to be useful?
- It's complicated! It's very dependent on the application / context. Big data is very abstract – what does it mean for me, for society?

**How could / should we work collectively to get greater impact?**

- Collectively, we need to commit to addressing the challenge of building robust understanding of big data, and to developing a shared framework and language to help.
- Getting people involved in data through citizen science, e.g. collecting and analysing data, including data from various sources
- Developing a more robust understanding of the public sphere and the potential for citizen participation (e.g. through the ideas of closed / uninvited spaces).
- Do we need a high level, cross-sector forum to deliver against these opportunities?

These and other points are picked up in the next section

# 3. Horizon scanning

In a horizon scanning exercise, delegates were presented with a list of suggested 'hot spots' and challenges in the areas of research and public engagement. Delegates reviewed these, and added additional topics which were agreed to represent both challenges and opportunities for public engagement. The key areas that were highlighted are recorded below.

The variety and quantity of topics identified was striking – far too long a list to act on. How should these be prioritised, and how might a coordinated approach to tackling them be realised given the current 'silo' working that was identified?

## Some emerging research areas / new applications of big data

- Confidence in the robustness of the decisions taken by machines, in tandem with increasing profundity in machine learning
- Factors affording successful interaction between people and computers
- Improving the transparency and interpretability of machine learning
- How systems cope with real world biases and 'messiness'
- The contribution of high resolution environmental data (e.g. air pollution sensors) to health and environment
- Exploring new approaches to protecting privacy, for example through using "synthetic data"
- How patient data could be linked to other data sources to provide greater insights into health and illness
- The merging of AI and Big Data – e.g. in hospitals
- The influence of Big Data on service delivery
- Real time dynamic and predictive diagnostics in medicine and all aspects of personal data

Other factors which will impinge on the research landscape were identified:

- Developing research methods that better integrate perspectives from beyond academia
- The need to ensure a safe and secure environment to play around with data
- Team science – we need a mix of skills to do and to communicate the research and engage with the public and will need to invest in this
- We need to consider the economic drivers of research in this area: as business takes the best talent (with big salaries), will university research fall behind?

## Challenges in understanding the dynamics of public engagement

- Trying to understand how humanity can manage innovation – how 'be human' in this rapidly changing context?
- How to improve digital literacy? Can we develop a common language to engage people? If so how should it be disseminated and embedded?
- How do attitudes change as technologies diffuse through society?
- How can we support consumers to better understand the business models of the large 'data controllers' or digital service providers?
- Public engagement should be exploring behaviour as well as attitudes
- Understanding new forms of responsibility and accountability
- How consent is managed in other fields (e.g. doctors seeking consent from patients): could learning here be applied to the data field?
- Do people understand how much data infuses their lives already?
- How aware are people of governance mechanisms and how to claim redress? Are organisations even explaining privacy and governance? So do we know cause and effect?
- There is lots of activity in the area of health data, but a deficit in other areas which should be addressed
- Understanding how publics view different data contexts and how these influence their attitudes and behaviour

Delegates were invited to identify some priority areas for future public engagement. The list below captures the key areas identified through the briefing paper and discussion at the event.

## Some emerging areas for debate and dialogue

- The appropriate governance of new uses of data**.** This is an area of intense activity currently and there have been a number of proposals for how the governance of this area might be improved. If any of these proposals are to move forward effectively they will both have to (i.) engage the public in their development, and (ii.) be structured in such a way that the public is able to engage substantively in the issues they are dealing with
- Work being led by the Farr Institute to develop a Consensus Statement for using data in research to launch later this year
- The impact of autonomous systems on employment and skills, and the appropriate distribution of the benefits arising
- Tracking how public attitudes evolve over time
- Tracking the effect (if any) of "hot topics" (like 'care.data') on the public's awareness and views of the collection and use of personal data
- Exploring public views on specific Big Data innovations. Public views on the collection, sharing and use of personal data can vary considerably depending upon the context. Other types of Big Data – e.g. climate data – may not raise the same privacy issues. It will be important to review public views on specific Big Data innovations to understand the nuance of public opinion in different contexts
- Exploring the factors which affect how the public makes trade-offs. The public approaches different trade-offs in different ways depending upon the data in question and the possible advantages and disadvantages. Exploring in more detail what factors affect how these trade-offs are made would be useful for anticipating the public's response to a particular scenario
- How we can best support the public to understand topics like how statistics work, what happens with their data and how data becomes combined into datasets
- Better understanding the nuances of public opinion in this area
- The role and contribution of citizen science
- The need to take account of the increasing inequality, and how this will impact on the data landscape
- Rather than always being reactive to the technology, can we be more proactive in exploring how society views the potential benefits?
- The potential of data to be exploited to generate misinformation and massive social manipulation – held in tension with its potential to support democratising of knowledge; citizenship; empowerment – there is significant positive potential which could be further explored
- Data determinism – people being 'labelled' and pigeon holed by their data
- Use of algorithms – for instance in insurance, recruitment, prison system: how can we ensure transparency and appropriate governance?
- Privacy enhancing technologies: these will have a significant impact on governance, and create significant business opportunities
- The increasing cost of data protection outweighing the cost of attack
- Digital personhood
- Trust and security
- Safeguarding data

# 4. What next?

In the final session, delegates were invited to identify key challenges in the area of public engagement with big data, where they felt concerted attention and activity should be invested. The 6 topics chosen are summarised below.

The groups addressing each challenge had a very short time to develop their thinking. It was generally agreed that this was a really useful exercise and that time and resource could usefully be invested by government in developing this approach to prioritised topics.

**Challenge? Big data for political purposes: fake news and political manipulation.**

**Why?** Fake news and propaganda have always been a problem, but big data is a tool for doing it in a more widespread way.

**What engagement?** To understand what they want from democracy and political engagement; who they trust.

**Success looks like?** Clearer understanding of what democracy and freedom looks like in a big data era; a formal and informal policing system implemented.

**Challenge? The impact of big data and related tech on jobs and the UK economy**

**Why?** Rise of techniques to replace or make redundant the skills of data related activities around the economy.

**What engagement?**
Are we ready to trust automated systems instead of people? Are we training people in expectation of this change?

**Success looks like?** UK a data-enabled economy where we are taking advantage / are prepared for the impact of these technologies

**Challenge? Personal data and consent for research**

**Why?** GDPR; variety of challenges around people's understanding of consent; role of networks and individuals

**What engagement?**
Build on existing activity eg Patient Data

**Success looks like?**

- People in charge of their own data
- Informed consent – boundaries and impact on network explicit
- Knowledgeable public; Evidence based policy

**Challenge? Data literacy**

**Why?** Technical, complex area with very low understanding. Need a starting point / foundation to explore public attitudes

**What engagement?**
Involve key people working in the area: schools / intermediaries – trusted professionals / Technology providers / Privacy groups and advocacy groups

**Success looks like?** Improved data literacy through e.g. campaigns, citizen science, balanced reporting, informed consent / decisions

**Challenge? Privacy and security enhancing technologies**

**Why?** A big opportunity exists to use technologies and business models to manage the social and ethical tensions between USING data and protecting sensitive information

**What engagement?** Public / consumer testing of techs and business models; build on existing trust frameworks and data dialogues

**Success looks like?** New economic models; Better sharing of health data for research and clinical use; Engagement between research, public and business; Greater willingness to share data due to the security the technologies provide

**Challenge? Algorithms and delivery of services**

**Why?** Algorithms will be increasingly used to address needs, identify gaps, deliver services

**What engagement?** Involve communities; draw on good examples in private sector (AXA)

**Success looks like?** Engaged public service delivery
Engagement of workforce

# Some concluding comments

The points below seek to distil the key points that emerged over the course of the workshop, and in the final plenary.

Do these accurately reflect the 'big ideas' that you took away from the workshop? What's missing that you think is significant to include?

Delegates argued that our collective efforts to engage the public with big data developments are based upon explicit **ethical principles**. Two principles were consistently referenced in discussions: integrity and social justice

**Guiding principles**

- Integrity – how can we be confident in the motives and practices of people working to research, apply and commercially exploit the potential of big data?
- Social justice – how can we take steps to ensure that the developments in big data impact positively on social justice?

A number of '**hot topics**' kept surfacing throughout the discussion: topics where it was felt that debate and dialogue should be focused.

There was a strong steer that each of these topics also required concerted efforts to better support **public understanding**.

**'Hot' topics that require discussion**

- Privacy - how can individuals right to privacy be protected?
- Consent – how can we ensure people understand what they are consenting to? And that the consent remains relevant in light of future technological developments?
- Provenance – understanding the 'data journey' and how consent follows data over time, and across different uses
- Ownership – who owns your data, and how can you protect your data property?
- Value – what is data worth? To whom? How is that value exploited? How are the gains shared?
- Decision making and taking – how is data being used to automate decisions, and what controls do we feel are necessary?

**Three big strategic challenges emerged** in this area: governance; social intelligence and education.

.

**How data is managed and used: strategic challenges**

- Governance – what systems do we need to ensure appropriate regulation and ethical oversight is in place and various societal interests are balanced?
- Social intelligence – how can we ensure that rich and accurate social intelligence is being gathered to inform the rapid advances in the area?
- Education – how can we ensure that we are equipping all citizens – from children through to professionals in key service roles – to make sense of and negotiate their identities in a data driven world?

There was significant enthusiasm about the way that the workshop brought together people working with **common purpose**, but in different domains. There was seen to be value in the attempt to capture the broad landscape of Big Data, to help accelerate and align activity.

**'Joined up' working**

- The big data landscape 'map' could provide a useful device to encourage better coordination and alignment between the multiple actors active in this space
- There was enthusiasm for a more coherent and strategic approach to addressing public understanding of big data; and to investing in 'big' and 'small' dialogues which build our collective knowledge base rapidly and coherently

The meeting ended with the BEIS Sciencewise team committing to bringing together key people from across government working on data issues, to review the findings from the workshop and to consider whether Sciencewise might support follow up activity.

Generally, we were struck by the enthusiasm expressed for some kind of high level, cross-sector stakeholder forum to build on the quality and focus of the conversations at the workshop.

We would welcome your comments on this draft report. In particular, we would like to invite you to submit your suggestions for:

- How BEIS / Sciencewise could act on the findings from the workshop
- How you / your organisation would like to contribute to any follow up
- Who's missing - what organisations which have an important role to play in these fields weren't involved in the workshop?
- What else you think should be done, by whom, to capitalise on the momentum

# Appendix: Briefing paper

Department for Business, Energy & Industrial Strategy

sciencewise EXPERT RESOURCE CENTRE

# Big Data workshop

## Briefing paper

Friday 22nd September, 10.30 – 3.30
BEIS Conference Centre, 1 Victoria St, Westminster, London SW1H 0ET

Thank you for agreeing to attend the workshop on the 22nd September.

The event is supported by the BEIS Sciencewise programme. The Sciencewise programme helps policy makers to deepen their understanding of the public's views on new and emerging areas of science and technology, through deliberative dialogue, enabling them to develop policies which take account of public opinion. The event is being run in partnership with the National Coordinating Centre for Public Engagement (NCCCPE). The NCCPE is funded by RCUK, Wellcome and the HE Funding Councils to increase the quality, coordination and impact of public engagement in the higher education sector.

We are delighted that we will be joined by such an expert and experienced group of people who are working in research, policy and practice, in a variety of roles, to realise the potential of big data. We hope that by working together in a really focused way we can stimulate purposeful shared learning, and more importantly, identify an agenda for future activity.

This paper provides a brief overview of our plans for the workshop. We hope you will have time to read through it in advance. We would welcome any comments in advance of the workshop. It contains a number of prompt questions for delegates:

- Does the representation of the field of Big Data research and application (page 5) provide a 'good enough' map to help us describe the complex big data landscape, as it is evolving in research, policy and practice? Where does the activity you are involved in fit on that map?
- Does our summary of the key lessons learned from the public engagement to date reflect your understanding? What is missing or significant which we need to take account of? What do you think are the strengths and weaknesses of the work done to date?
- Looking ahead from your perspective, what do you anticipate being the most significant developments in research and application in the field of Big Data? What are the implications of these for future public engagement?
- Are there specific actions which you think BEIS through the Sciencewise programme and/or other agencies should be taking to future proof public engagement in this area?

**Simon Burall - Programme Director, Sciencewise**
**Paul Manners - Director, National Coordinating Centre for Public Engagement**

# Attending

We are delighted that the following individuals and organisations will be represented at the event.

## Policy and practice

- Nicholas Dodd – Department for Digital, Culture, Media and Sport
- Susan Krouwel - Department for Digital, Culture, Media and Sport
- Marcus Besley – Go-Science
- Nicola Perrin – Head of Policy, Wellcome
- Edward Blandford – Senior Science Officer, Department of Health
- Sue Bateman – Head of Policy, Transparency Team, Cabinet Office Cabinet Office
- Jenni Chambers – Head of Public Engagement with Research, RCUK
- Simon Gardner – Joint Head of Innovation Programmes and Partnerships, Science and Innovation Directorate, NERC
- Lucy Geoghegan – Industrial Strategy, Department for Business, Energy and Industrial Strategy
- Laura Riley – Head of Ethics, Genomics England
- Natasha McCarthy – Head of Policy (data), Royal Society
- Katie Weekes, Public Engagement Manger, Royal Society
- Tim McGarr – Market Development Manager for Information Technology, British Standards Institute
- Philippa Westbury – Policy Adviser, Royal Academy of Engineering
- Hetan Shah – Executive Director, Royal Statistical Society
- Simon Briscoe – Vice Chair UK Data Service and Deputy Chair of ESRC Data Infrastructure Strategic Advisory Committee
- Imogen Parker – Programme Head, Public Administration and Law, Nuffield Foundation
- Vince Smith, Head of Informatics, Natural History Museum
- Helena Quinn – Policy Officer, Alan Turing Institute
- Sophie McIvor – Head of Communications, Alan Turing Institute

## Researchers

- Prof Vania Sena - Head of the Management Science and Entrepreneurship group at Essex Business School and director of the ESRC Business and Local Government Data Research Centre
- Prof Peter Smith - Professor of Social Statistics at the University of Southampton and Director of the Administrative Data Research Network and Director of the Administrative Data Research Centre for England.
- Keith Dingwall – Senior Business Manager, Urban Big Data Centre, University of Glasgow
- Prof Sofia Olhede – Director, UCL Centre for Data Science
- Louise Corti - Associate Director, UK Data Archive at University of Essex
- Prof Mark Birkin – Director and Principal Investigator, UK Consumer Data Research Centre
- James Wilson – Vice-Dean (Interdisciplinarity) for the Faculty of Arts and Humanities, and co-director of the UCL Health Humanities Centre
- Prof Lorna McGregor - Principal Investigator and Co-Director of ESRC Human Rights, Big Data and Technology grant, University of Essex
- Prof Glenn Parry – Professor of Strategy and Operations Management, UWE Bristol
- Adam Rae - Head of Urban Data, Future Cities Catapult
- Simon Jude - Cranfield Institute for Resilient Futures and Data, Risk & Environmental Analytical Methods (DREAM) Centre for Doctoral Training (TBC)
- Professor Sarah Cunningham-Burley – Dean of Molecular, Genetic and Population Health Sciences, University of Edinburgh and Head of Public Engagement, Farr Institute
- Jacky Pallas - Director of eResearch at Kings College London

## Facilitators and hosts

- Paul Manners – Director, National Coordinating Centre for Public Engagement
- Simon Burall – Programme Director, Sciencewise
- Hayley Gowen - Public Engagement in Science and STEM Inspiration, BEIS
- Tony Whitney - Public Engagement in Science and STEM Inspiration, BEIS
- Alec Weir - Public Engagement in Science and STEM Inspiration, BEIS

# The focus of the workshop

We plan to work through the following steps in the workshop

## 1. Agree a shared representation of the 'field' of big data research

We will present a simple mapping of the field (see page 5), which we plan to review and improve at the workshop.

Our goal is to ensure that we are all 'on the same page' in describing what we understand by Big Data and how it is currently being researched, and that research applied.

## 2. Take stock of the public engagement that has been undertaken to date and the key lessons learned

There has been a variety of public engagement and public dialogue projects, some from the early 2000s.

We have developed a synthesis of the key insights gleaned from these activities (pages 6 and 7), which will present and review with your help.

We want to identify and test what we think we already know, and to identify areas where our current knowledge is limited.

## 3. To horizon scan to identify critical 'hotspots' in research and application over the next 5 years – the most significant likely breakthroughs, and their implications for policy making and wider social application, and for public engagement

Having reviewed the research and public engagement activity that has been underway, we will then start looking ahead.

We want your help to explore how you anticipate the field developing in the next five years.

From your perspective, what do you anticipate being the most significant developments in research and application?

What are the implications of these for public engagement?

## 4. To identify a strategic response to the above, agreeing how future public engagement activity might be better targeted and prioritised to address the opportunities and risks identified

The last part of the workshop will focus on what needs to be done to ensure that we are investing in appropriate forms of public engagement in this area, and coordinating this activity effectively.

## 1. Agree a shared representation of the 'field' of big data research

This table attempts to describe the terrain, and the key features of the landscape. It is of course a significant simplification.

We will invite you to review this and to identify where the work you are involved in fits (or doesn't).

**Does this provide a 'good enough' map to help us describe the complex big data landscape, as it is evolving in research, policy and practice?**

**Defining Big Data**

We are using the definition of Big Data offered in the recent Royal Society and British Academy Report 'Data Management and Use: Governance in the 21st Century' (2017)

*'Large and heterogeneous forms of data that have been collected without strict experimental design. Big data is becoming more common due to the proliferation of digital storage, the greater ease of acquisition of data (e.g. through mobile phones) and the higher degree of interconnection between our devices (i.e. the internet).'*

| Domain | Focal points for investment and activity | | |
|---|---|---|---|
| Big Data **infrastructure**: improving the dynamics of data creation, access, storage and use | • Technical infrastructure<br>• Interoperability and access<br>• Analytics<br>• Capability and capacity | | |
| Big Data **categories:** the types of data being researched and applied | • Archive data<br>• Environment data<br>• Research data<br>• Clinical / health data<br>• Longitudinal survey data | • Admin data<br>• Business data<br>• Open data (public sector)<br>• Social media data | |
| Big Data **research domains:** areas of society in which applications of data are being researched | • Health and social care<br>• Services<br>• Government<br>• Energy | • Environment<br>• Creative economy<br>• Business<br>• Transport | • Cities<br>• Civil society |
| Big Data **applications**: areas where data has transformational potential | • Machine learning (including autonomous vehicles)<br>• The Internet of Things<br>• Artificial and augmented intelligence<br>• Decision making and service design | | |
| Big Data **research themes**: focal points for investigation | • Increasing the accessibility and usability of the data<br>• Trust, identity, privacy and security: the ethical dimensions to data capture and use<br>• Social and cultural applications: how data can transform everyday life, from culture to health<br>• Commercial application: exploiting the potential of data to unlock economic benefit | | |
| Big data **stakeholder groups**: the key types of organisations with a stake in the area | • **Research**<br>  • *Research funders*<br>  • *Learned societies*<br>  • *Research leaders / principal investigators* | • **Policy making**<br>  • *National government departments*<br>  • *Other government / policy agencies*<br>• **Societal stakeholder groups**<br>  • *'Watchdogs'*<br>  • *Societal stakeholders*<br>  • *Professional representation* | |

## 2. Take stock of the public engagement that has been undertaken to date and the key lessons learned

At the workshop we will present a synthesis and summary of key developments in public engagement in the Big Data field, and invite you to reflect on the following questions

### Social intelligence

**Do we have the mechanisms in place to ensure we are capturing and sharing effectively intelligence about public views, attitudes and understandings?**

There have been numerous public dialogues since 2002. A Sciencewise briefing report in 2014 reviewed many of these; the Understanding Patient Data website also contains a useful summary and also identifies examples of good practice in public and patient engagement. Some key themes arising from these reviews are identified on the next page.

### Engagement activity

**Are we investing in a judicious mix of high quality engagement activities to maintain open, two-way interaction with the public?**

We do not currently have a comprehensive picture of the activity that is underway, or of its quality and impact. The Understanding Patient Data website contains principles of good practice; but currently there is no coordinated approach to commissioning public engagement activity or pooling expertise. Should there be?

### Engagement capability

**Do we have the skills and capability to commission, undertake and evaluate effective public engagement, in the policy, practice and research communities?**

There are pockets of expertise (e.g. the Administrative Data Research Centre has a public engagement team). Work on Open Policy making and public engagement in higher education is beginning to address some of the cultural factors which inhibit public engagement. This is work in progress.

## 2. Take stock of the public engagement that has been undertaken to date and the key lessons learned

### Reviewing what we have learned from the public engagement already undertaken

**The struggle to make sense**
This is a complex and often abstruse area – finding robust but simple ways to scaffold understanding is a challenge.

**Differing perspectives**
Various 'segmentations' have been attempted to describe the noticeable differences in how different people perceive the area. For example, the Ipsos Mori Public Dialogue on the Ethics of Data Science in government identified these four clusters:

- '**data adopters**' (23% of adults) who support using data science for research purposes and see the value in how individual level data can generate better insight;
- '**data adapters**' (28% of adults) who respond best to uses which improve services for individuals and use of non-sensitive data;
- '**data pragmatists**' (27% of adults) who are more ambivalent in their views, wanting government to explore new ways of using data but are most comfortable using data for high-level statistics rather than advanced data science; and
- those who are '**data wary**' (22% of adults), who apply caution to the principle of data science, based on concerns around privacy and effectiveness or a desire for further information.

The 'One Way Mirror: Public Attitudes to Commercial Access to Health Data' identified 7 different mind-sets, represented in figure one.

**Red lines**
Consistently, certain 'red lines' emerge: areas where there is heightened public concern. These are usefully represented in the 'One Way Mirror' report (see figure 2).

- Consistently, we have learned that WHO is collecting and using the data (and their assumed motives) will influence how the public react. Businesses, particularly the insurance industry, are viewed with suspicion.
- The overriding question for everyone is the WHY. If there is a public benefit, they will be less concerned about the WHO.
- There is deep anxiety about privacy and security.
- There is concern about 'private' benefit. Clearly articulated public benefit is important for people to feel comfortable about new developments.

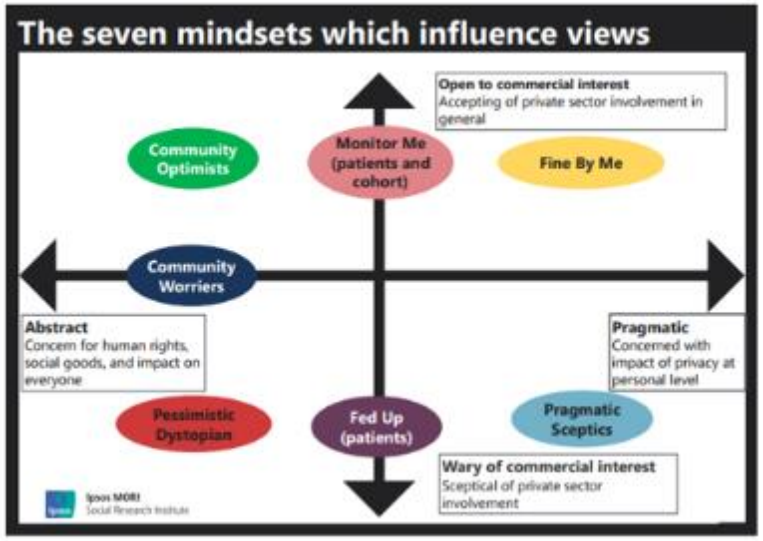


Figure one

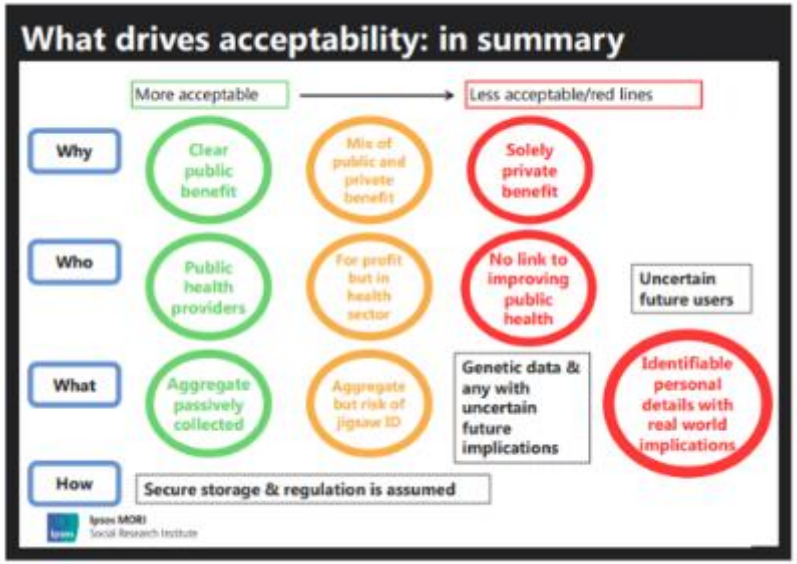


Figure two

**3. To horizon scan to identify critical 'hotspots' in research and application over the next 5 years – the most significant likely breakthroughs, and their implications for policy making and wider social application, and for public engagement**

Our review of the various reports and activity in the area of Big Data has helped us to identify the following areas where people have identified 'hot spots' or challenges which need attention. We will review and add to these at the workshop, and take stock of the current activity which we are aware has been commissioned.

- **What do you think are the most significant emerging areas of research and application in the next five years?**
- **What are the implications of these for public engagement?**
- **Are the plans that we know to be in place adequate to the challenges that these areas will pose?**

## Some emerging research areas

- Confidence in the robustness of the decisions taken by machines
- Factors affording successful interaction between people and computers
- Improving the transparency and interpretability of machine learning
- How systems cope with real world biases and 'messiness'
- The contribution of high resolution environmental data (e.g. air pollution sensors) to health and environment
- Exploring new approaches to protecting privacy, for example through using "synthetic data"
- How patient data could be linked to other data sources to provide greater insights into health and illness

**Please add to this list**

## Some emerging areas for debate and dialogue

- The appropriate governance of new uses of data**.** This is an area of intense activity currently and there have been a number of proposals for how the governance of this area might be improved. If any of these proposals are to move forward effectively they will both have to (i.) engage the public in their development, and (ii.) be structured in such a way that the public is able to engage substantively in the issues they are dealing with
- Work being led by the Farr Institute to develop a Consensus Statement for using data in research to launch later this year
- The impact of autonomous systems on employment and skills, and the appropriate distribution of the benefits arising
- Tracking how public attitudes evolve over time
- Tracking the effect (if any) of "hot topics" (like 'care.data') on the public's awareness and views of the collection and use of personal data
- Exploring public views on specific Big Data innovations. Public views on the collection, sharing and use of personal data can vary considerably depending upon the context. Other types of Big Data – e.g. climate data – do not raise the same privacy issues. It will be important to review public views on specific Big Data innovations to understand the nuance of public opinion in different contexts
- Exploring the factors which affect how the public makes trade-offs. The public approaches different trade-offs in different ways depending upon the data in question and the possible advantages and disadvantages. Exploring in more detail what factors affect how these trade-offs are made would be useful for anticipating the public's response to a particular scenario
- How we can best support the public to understand topics like how statistics work, what happens with their data and how data becomes combined into datasets
- Better understanding the nuances of public opinion in this area
- The role and contribution of citizen science

**Please add to this list**

## 4. To identify a strategic response to the above, agreeing how future public engagement activity might be better targeted and prioritised to address the opportunities and risks identified

The final part of the workshop will allow us to take stock, and to identify concrete areas where we believe, collectively, that concerted activity is necessary.

We will start with some quite broad questions to allow us to take stock of the territory:

- **What is our view of the current 'health' and robustness of the public engagement activity that has been undertaken and is planned?**
- **How can the strengths be built upon? How can any weaknesses be addressed?**
- **Are we exploiting effectively enough the existing knowledge and insight that has been gathered?**

We will then move to identify specific interventions which we believe are important, and seek to define a rationale for any recommended next steps:

- **Are there areas where we believe further public dialogues need to be commissioned?**
- **What more needs to be done to strengthen the governance of this area?**
- **What other investments would be wise, and why?**

BEIS is working with Involve, experts in public dialogue, and with the NCCPE to run the Sciencewise programme. The programme offers expert advice and support to plan and run a public dialogue, including a framework on contractors experienced in leading dialogues on complex issues and their evaluation. BEIS can fund up to 50% of a public dialogue led by a government department or agency.

If you have comments or queries about this briefing paper, or the workshop, please contact Paul Manners at the NCCPE:
paul.manners@uwe.ac.uk
07795 288133