



Department for  
Business, Energy  
& Industrial Strategy



# Big Data workshop

Briefing paper

Friday 22<sup>nd</sup> 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 22<sup>nd</sup> 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**

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 Mclvor – 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 21<sup>st</sup> 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 <b>infrastructure</b> : improving the dynamics of data creation, access, storage and use	<ul style="list-style-type: none"> <li>• Technical infrastructure</li> <li>• Interoperability and access</li> <li>• Analytics</li> <li>• Capability and capacity</li> </ul>		
Big Data <b>categories</b> : the types of data being researched and applied	<ul style="list-style-type: none"> <li>• Archive data</li> <li>• Environment data</li> <li>• Research data</li> <li>• Clinical / health data</li> <li>• Longitudinal survey data</li> </ul>	<ul style="list-style-type: none"> <li>• Admin data</li> <li>• Business data</li> <li>• Open data (public sector)</li> <li>• Social media data</li> </ul>	
Big Data <b>research domains</b> : areas of society in which applications of data are being researched	<ul style="list-style-type: none"> <li>• Health and social care</li> <li>• Services</li> <li>• Government</li> <li>• Energy</li> </ul>	<ul style="list-style-type: none"> <li>• Environment</li> <li>• Creative economy</li> <li>• Business</li> <li>• Transport</li> </ul>	<ul style="list-style-type: none"> <li>• Cities</li> <li>• Civil society</li> </ul>
Big Data <b>applications</b> : areas where data has transformational potential	<ul style="list-style-type: none"> <li>• Machine learning (including autonomous vehicles)</li> <li>• The Internet of Things</li> <li>• Artificial and augmented intelligence</li> <li>• Decision making and service design</li> </ul>		
Big Data <b>research themes</b> : focal points for investigation	<ul style="list-style-type: none"> <li>• Increasing the accessibility and usability of the data</li> <li>• Trust, identity, privacy and security: the ethical dimensions to data capture and use</li> <li>• Social and cultural applications: how data can transform everyday life, from culture to health</li> <li>• Commercial application: exploiting the potential of data to unlock economic benefit</li> </ul>		
Big data <b>stakeholder groups</b> : the key types of organisations with a stake in the area	<ul style="list-style-type: none"> <li>• <b>Research</b> <ul style="list-style-type: none"> <li>• Research funders</li> <li>• Learned societies</li> <li>• Research leaders / principal investigators</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• <b>Policy making</b> <ul style="list-style-type: none"> <li>• National government departments</li> <li>• Other government / policy agencies</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• <b>Societal stakeholder groups</b> <ul style="list-style-type: none"> <li>• 'Watchdogs'</li> <li>• Societal stakeholders</li> <li>• Professional representation</li> </ul> </li> </ul>

## 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.

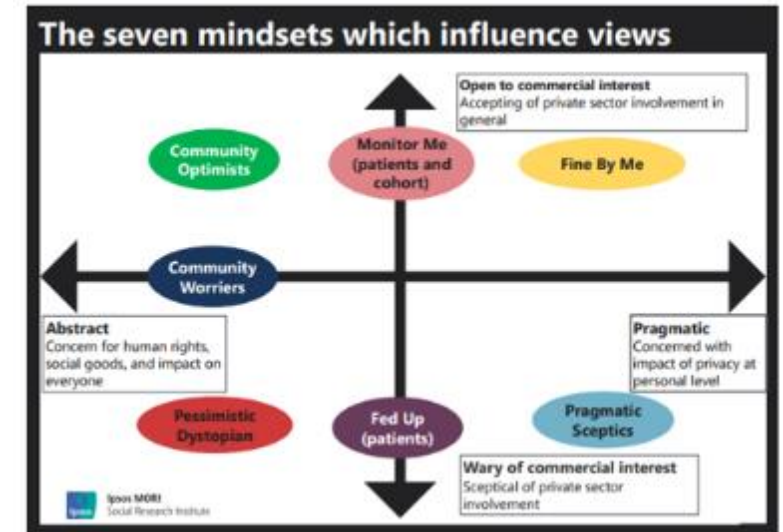


Figure one

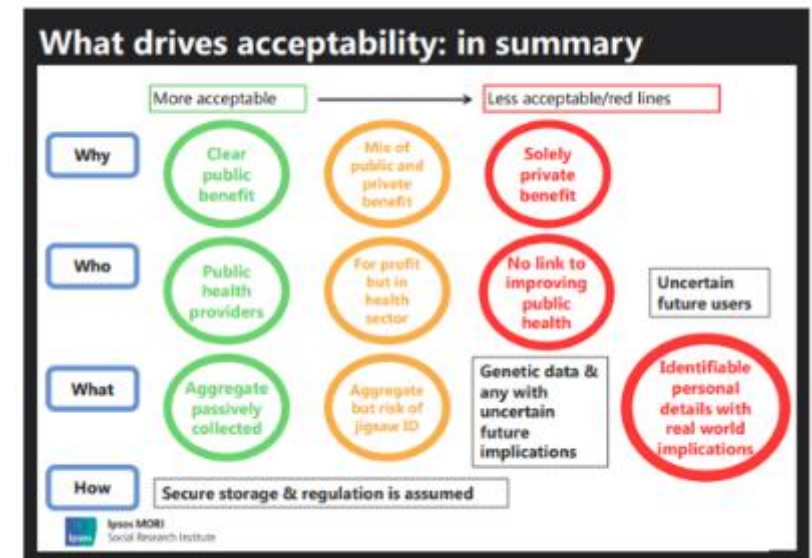


Figure two



### 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](mailto:paul.manners@uwe.ac.uk)

07795 288133