



This is the author's version of a work that was accepted for publication:

Michael, M., & Lupton, D. (2016). Toward a manifesto for the 'public understanding of big data'. *Public Understanding of Science*, 25(1), 104-116.

<https://doi.org/10.1177/0963662515609005>

This file was downloaded from:

<https://researchprofiles.canberra.edu.au/en/publications/toward-a-manifesto-for-the-public-understanding-of-big-data>

©2015 Mike Michael & Deborah Lupton

Notice: This is the authors' peer reviewed version of a work that was accepted for publication in *Public Understanding of Science*, which has been published at <http://doi.org/10.1177/0963662515609005>.

Changes resulting from the publishing process may not be reflected in this document.

Toward a Manifesto for the ‘Public Understanding of Big Data’

Mike Michael, University of Sydney
Deborah Lupton, University of Canberra

This is the accepted author’s version (postprint) of an article published in *Public Understanding of Science*, 2016, 25(10), 104–116: <http://pus.sagepub.com/content/25/1/104>

ABSTRACT

In this paper, we sketch a ‘manifesto’ for the ‘public understanding of big data’. On the one hand, this entails such PUS and PEST-tinged questions as: How, when and where are people exposed to, or engage with, big data? Who are regarded as big data’s trustworthy sources, or credible commentators and critics? What are the mechanisms by which big data systems are opened to public scrutiny? On the other hand, big data generates many challenges for PUS and PEST: How do we address publics that are simultaneously the informant, the informed, and the information of big data? What counts as understanding of, or engagement with, big data, when big data themselves are multiplying, fluid and recursive? As part of our ‘manifesto’ we propose a range of empirical, conceptual and methodological exhortations. We also provide an annexe that outlines three novel methods for addressing some of the issues raised the paper.

Toward a Manifesto for the 'Public Understanding of Big Data'

Introduction

In this paper, we begin to sketch out a 'manifesto' of what it means to engage with publics and their relations to the complex issues thrown up by 'big data'. The term 'big data' is used to describe the massive and continually-generated digital datasets that are produced via interactions with online technologies. These interactions may involve both human and nonhuman actors, particularly as the 'Internet of Things' and 'sensor society' develop. Human actors contribute to big datasets when they engage in such activities as making calls and using apps on mobile phones, using online search engines such as Google, purchasing goods or services online or taking part in customer loyalty programmes, uploading contributions to social media platforms, using wearable self-tracking devices or moving around in spaces that are equipped with digital sensing or recording devices. Nonhuman actors participate in the generation of digital data via such technologies as self-tracking devices, 'smart agriculture', in which plants and animals are fitted with sensors, the 'smart city', which includes such technologies as CCTV cameras and sensor-based traffic management systems and the 'smart home', involving domestic appliances producing and sharing data on the habits and movements of their occupants.

A digital data knowledge economy has developed in recent years, in which digital data have taken on commercial, research and managerial value for many actors and agencies. As the aggregation of individuals' personal (or 'small') data into massive data sets, big data are represented as offering unprecedented insights into social and economic trends and behaviours. Big data in their multiple manifestations have also become a 'big issue' for the social sciences. Big data is both a topic (what are the social impacts of big data?) and a resource (how does the use of big data challenge existing sociological accounts of social processes, and how might the social sciences respond to, even learn from, this?). The complexity is neatly summed up in an introductory statement on the home page of the new interdisciplinary journal *Big Data and Society*:

The Journal's key purpose is to provide a *space* for connecting debates about the emerging field of Big Data *practices* and how they are reconfiguring academic, social, industry, business and government relations, expertise, methods, concepts and knowledge. (*Big Data & Society*, 2014, emphasis in the original)

Taking a lead from this statement, the present paper attempts to connect big data specifically to debates within the broad multidisciplinary fields of public understanding of science (PUS) and public engagement with science and technology (PEST). *Prima facie*, big data lends itself to PUS and PEST framings of 'the' public. After all, it can be said to represent a technically specialist set of practices by which data about, for example, people's 'online' behaviour is gathered, processed and represented. In this respect, to focus on big data is to ask such PUS

and PEST-tinged questions as: How, when and where are people exposed to, or begin to engage with, big data? Who are regarded as trustworthy sources of big data, and credible commentators upon, and critics of, them? How do big data reflect, mediate and respond to folk knowledges about, for instance, the use and misuse of information? What are the mechanisms by which big data systems and their development are opened to public scrutiny, or which allow for the impact of publics' voices?

On the other hand, big data also raise a series of issues that can illuminate the presuppositions built into PUS and PEST framings. As we shall see, big data take on many forms, and their relation to publics is highly variegated, being at once a description of, a resource for, and a product that depends on, public practices. Such practices can range from 'active' (for example, the use of social media) to 'passive' (for example, where surveillance technologies gather data). In any case, big data map onto debates in PUS and PEST about how publics are partially constituted by the expert practices entailed in the empirical and analytic machinery of PUS and PEST (for example Irwin and Michael, 2003). In particular, the 'involutions' of big data in which publics are, in varying degrees, both the subjects and objects of knowledge, both authors and texts, simultaneously informants, information and informed, suggest that PUS and PEST might need to expand their empirical and analytic repertoires if they are to do justice to the phenomena of big data.

In what follows, we begin with a brief overview of some of the key debates in PUS and PEST. We then present an account of the various concerns associated with big data. We bring these together and ask how PUS and PEST might respond to these issues, conceptually, empirically and methodologically. In the process, we begin to sketch a 'manifesto' for a 'public understanding of big data' (PUBD – see below). By way of proviso, this is self-evidently an exploratory piece, and the goals and strategies that comprise our manifesto are presented circumspectly. Nevertheless, our modest hope is that this paper prompts discussion about the ways in which public understanding of science/public engagement with science might respond to the complex challenges posed by big data.

PUS/PEST

In the earliest studies of science-society relations quantitative techniques such as surveys were used to measure lay people's scientific literacy, or to gauge their deficiencies in scientific knowledge. With the rise of critical PUS, the so-called deficit model (namely, the *de facto* focus on the public's scientific ignorance) was abandoned and emphasis placed on the qualitative analysis of lay people's folk knowledges, which included case studies of their trust in particular scientific institutions. Here, qualitative methods such as ethnographic observation, semi-structured interviews and focus groups accessed the complex ways in which lay people's knowledge was bound up with issues of identity, relations of power, and tacit skill. If quantitative PUS was shaped by an interest in facilitating scientific understanding in order to encourage public support of science, qualitative (or critical PUS) paid attention to the ways in which lay people's existing knowledges were oftentimes denigrated by scientific experts

with the result that the credibility of scientific institutions become progressively eroded (for a classic overview, see Wynne, 1995).

Subsequently, there have been a number of developments in the field. For example, the divide between expert and public constituencies – science and society – has been reconfigured. Thus, for instance, the Mode II model of science (Nowotny et al., 2001) suggests that science and society increasingly interdigitate as lay people come to be situated as lay experts or scientific citizens. These new configurations have been variously theorized as, for instance, assemblages or hybrid forums (see Irwin and Michael, 2003; Callon et al., 2001). The interdigitation of science and society became also a proactive project within the field. This, of course, reflected the ‘participatory turn’ in which ‘public understanding of science’ transitioned into ‘public engagement with science and technology’. Crucially, PEST was concerned with developing and testing deliberative techniques by which the voice of the public could enter into decision-making processes around science and technology (Hagendijk and Irwin, 2006).

A range of deliberative or participatory methods have emerged (for example citizens’ juries, consensus conferences, deliberative polling, card-based group discussion), and PEST has been centrally focused on how to ensure their ‘appropriate functioning’ (for example Chilvers, 2008). Further, PEST scholarship has interrogated the problematic assumptions built into these engagement procedures and the broader governmental functions that attach to them (Michael, 2009; Welsh and Wynne, 2013). This critical account articulates with recent pragmatist perspectives on publics, democracy and the emergence of political issues. As Marres (2012) has argued, PEST has been primarily focused – technically and critically – on the ‘procedural’ elements of the relations between publics and science. By contrast, she advocates a view that focuses on how issues – especially those that escape institutional framing – co-emerge along with their publics. Thus Marres suggests that publics can articulate issues that have failed to register institutionally. Accordingly it is in the process of that issue articulation that the public comes into being. Of course, this is not to say that these ‘issues in the wild’ do not inter-relate with more procedural approaches: after all, the latter feed ‘imaginaries of publics’ (Welsh and Wynne, 2013) back to emergent publics, and emerging issues can be taken up as PEST initiatives that are institutionally and practically expanded (for example Hinchliffe et al., 2014).

It is within this evolving landscape of PUS and PEST that we wish to situate our Manifesto for a Public Understanding of Big Data (PUBD). At this point a word of explanation is in order for our use of the term ‘Public *Understanding of Big Data*’ (especially in light of the preceding condensed history of the field). Our main rationale for this is the intuition that in the case of big data, understanding and engagement become difficult to differentiate. To ‘misunderstand’ big data is no longer a matter of compromised literacy but, within the dynamics of big data, another opportunity for ‘engagement’. For instance, for a public to ‘misunderstand’ big data is, potentially at least, to contribute to yet more big data about such ‘misunderstanding’ which might subsequently be used to market something to – engage with – that ‘misunderstanding’ public. Public *Understanding of Big Data* thus simply serves as a heuristic term which ideally would serve to prime questions about the character of understanding, but also the complexity of ‘engagement’ (whether that be epistemic, political, cultural or economic).

Critical Data Studies

If the preceding section placed PUBD in relation to the concerns of PUS/PEST, the present section sets out to present a number of critical accounts of big data. This is by no means exhaustive, but simply indicative of the range of big data and the array of issues that have followed in their wake. This will serve as the basis of more detailed discussion of a Manifesto for PUBD.

A body of literature has begun to develop that may be loosely described as critical data studies (Kitchin & Lauriault, 2014) in response to big data. Critical scholars have investigated the ways in which big data are configured via human decision-making and the role played by the internet corporations and other digital developers in shaping how big data are gathered and analysed. Big data are therefore positioned in this literature as a sociotechnical assemblage involving complex interactions of humans and nonhumans (boyd & Crawford, 2012; Lyon & Bauman, 2013; Kitchin, 2014; Kitchin & Lauriault, 2014;).

The supposed objectivity and neutrality of big data have been challenged by writers in this literature. Critical scholars have called into question the notion of 'raw data' that underpins dominant representations of big data, pointing to the ways in which data are inevitably always partial, selective and 'cooked' by social actors. The conventions and practices of seeking out, recording, archiving and categorising data have been examined as part of this critique (Baym, 2013; Boellstorff, 2013; Gitelman, 2013; Räsänen & Nyce, 2013). Additionally, Boellstorff's (2013) concept of 'rotted data' draws attention to the impurity of these data, their materiality and the loss of control their generators have once the data enter circulation in the digital data economy. It has been further argued that big data are co-authored by the heterogeneous actors involved in their generation, including members of the public as well as a heterogeneous panoply of other actors such as code and software, platforms, sensors, programming and data analysis experts and the algorithms that are employed in selecting and making sense of big data (Beer, 2013; Beer & Burrows, 2013; Rogers, 2013; Totaro & Ninno, 2014).

A digital data assemblage is generated each time a user interacts with a digital technology that is incorporated into the internet and thus feeds data into a larger archive to be aggregated with others. Here the user is both producer and consumer of content, as denoted by the term 'prosumer' that is used in media and internet studies to describe the interactions of users with online technologies (Ritzer, 2014). Importantly, these data assemblages are responsive, dynamic and lively, constantly reconfigured as new data are generated and datasets are combined in different ways (Lash, 2006; Andrejevic, 2013; Beer, 2013; Beer & Burrows, 2013; Lyon & Bauman, 2013; Kitchin, 2014). In some cases, this process of data reconfiguration can take on more dramatic forms as in the case of data hacking and on-going revisioning of open source data or programs. This suggests that to the processes of consumption/production, we can add 'produsage', and accompanying the figure of the 'prosumer' (and its relation to data as 'products') is the 'produser' (and its relation to data as 'process' – see Bruns, 2009).

When interacting with online technologies, people are encouraged to view themselves as data-generating subjects or as groups of data subjects (Ruppert, 2011), and to interact with and interpret their data assemblages.

Data assemblages often participate in a cybernetic process. They both draw on individuals' personal data and contribute to large digital datasets, and in turn may be used to feedback to individuals (either as their own personal data or as the insights offered by aggregated data) (Beer & Burrows, 2013; Lupton, in press). For example, the Glow menstruation, ovulation and fertility monitoring app and platform for women encourages users to enter manifold aspects of their daily bodily functions and activities, including signs of ovulation, menstruation, mood, body weight, diet and exercise and whether conception has occurred. Most of these data must be manually entered by users, but some can be automatically generated by other devices that sync to the app, including wearable diet and activity trackers such as MyFitnessPal. As the apps' algorithms learn about the bodies of its users, they are able to make customized predictions about ovulation, menstruation and fertility. The users can view graphs and other data imagery that represent their own data to themselves, helping them identify patterns and relationships between their data and to forecast ovulation to avoid or facilitate a pregnancy. However their personal data are also automatically uploaded to the developers' digital archive and are used by the developers to generate insights into their users' fertility and conception patterns across a population of tens of thousands of users (it is claimed on the Glow website that the app has helped over 20,000 women conceive in the first year in which it has been available). These aggregated data are then fed back into the app's function and the information it provides to users, interacting with the user's personal data as part of the algorithmic predictions that it makes for each user.

The use of big datasets in surveillance activities ('dataveillance') is also relevant here (Best, 2010; Lyon, 2014). Over the past two years or so, there has been considerable news media reporting of the Snowden-leaked documents about the ways in which people's personal digital data and metadata have been surveilled by national security agencies in the US, UK and Australia. Taken together with accounts of various hacker attacks on data bases containing people's often very private and personal information (such as medical data or the information they have uploaded to adult dating sites), this has demonstrated to the public just how insecure these data are. Other controversial exposés of the ways in which people's personal data are manipulated by social media companies (for example the Facebook emotion manipulation project), or harvested and sold for commercial benefit to third parties by data brokers, have publicized the potential drawbacks of becoming a data subject. However there are still manifold ways in which personal information is collected and repurposed as part of big data sets of which publics may be unaware. Given the rapidly multiplying sites and spaces in which people are digitally surveilled (literally from the inside-out in the case of such objects as ingestible tablets equipped with microchips and the continuous digital glucose monitors used by some diabetics) and the ways in which the data sets they generate can be joined up, the extent to which members of the public are able to maintain awareness of how their bodies and selves are being digitally recorded and rendered into big data is questionable. Further, as the monitoring of individuals' bodies, energy use, work productivity, moods, social

relationships, purchasing habits, driving practices and so on becomes more routinized and widespread, the extent to which the subjects of this tracking can opt out becomes limited. People may have few choices about whether or not to participate as data-generating subjects.

As digital device users continue to surveill each other (for example on social media sites), as both public and private spaces become embedded with digital sensors that monitor increasingly detailed aspects of behaviours and movements of users of these spaces, and as citizens are increasingly expected to use digital devices to monitor their own behaviours, it has become impossible to avoid becoming a subject of dataveillance (Lyon & Bauman, 2013; Lupton, in press;). Given the dispersed, constantly moving and continual nature of dataveillance, or what Lyon and Bauman refer to as ‘liquid surveillance’, we are all ‘citizen sensors’ (Gabrys, 2014) by default, whether we agree to this role or not.

Big Data and the Public

Above, we hinted that as we move toward a manifesto for PUBD, terms which are commonly employed in PUS and PEST become complicated. We noted the critical approaches in PUS/PEST have drawn attention to the ways in which ‘the public’ as a phenomenon and object of research is configured by the actors involved in researching PUS/PEST. In relation to big data, ‘the public’ becomes even more problematic a concept, given the multiple shape-shifting in which different data subjects and objects participate. As a consequence, the ‘public’ of big data is a constantly moving virtual artefact that has varying meanings and constituents depending on which actors are seeking to define this entity and at which point in time they are seeking to do so.

‘Understanding’ or ‘knowledge’, as initially expressed in the PUS literature, whether expert or lay, has tended to be represented as a static possession. With critical PUS and PEST this has changed somewhat, and these concepts have been tied to a multiplicity of concerns, including trust, identity, dialogue and so on. However the responsive, dynamic, lively nature of the knowledges produced from online technologies changes the stakes somewhat. What exactly counts as knowledge or understanding where the data themselves are continually in flux for both lay people and experts? Moreover, knowledges are invested with different forms of value in the digital knowledge economy. Unlike most forms of lay knowledges that PUS and PEST researchers have investigated (as opposed to, for example, work on the labour and property rights of such ‘publics’ as tissue donors – eg Cooper and Waldby, 2014), as we noted above the knowledges that are generated and consumed by the public as part of digital datasets have direct commercial and other forms of value. Once the data enter the digital knowledge economy, those who generate the data can lose control of them as they are repurposed by others (Beer, 2013). This raises issues around the relations of knowledge to ‘ownership’ (which might often take opaque forms), and further nuances the figure of ‘prosumer’ as it engages with big data as ‘property’.

The notion of ‘expertise’ can be similarly interrogated in relation to the digital knowledge economy. Take the example of the practice of citizen-led science initiatives that are being reworked via digital data practices. New forms of ‘citizen science’ or ‘citizen sensing’ are part

of crowdsourcing efforts that equip people with digital devices to collect scientific data and upload these to be aggregated with others' data. The concept of the 'smart city' or 'sentient city' (Thrift, 2014) incorporates the intersection of various sensor-equipped nodes, including human and nonhuman actors, combining data from various standpoints to assist with efforts towards environmental sustainability. These initiatives have in turn altered notions of what is considered ideal 'citizenship' around data practices. To be engaged as an active and responsible citizen may come to include the willingness to use sensors and other digital devices to monitor one's own environmental footprint and the environmental conditions in one's neighbourhood and share these data with others (Gabrys, 2014). Just as we may speak of 'sentient cities', we may also refer to the 'sentient citizen' - a digital data-emitting node in the Internet of Things. This has implications for the concept of expertise. 'Expertise' becomes distributed among the humans and nonhumans actors that comprise sentient landscapes in which data are co-generated.

It is important to recognize that the liquidities and flows of big data are not necessarily unimpeded: blockages and resistances are part of the digital data economy. The scenario of the active citizen sensor, for example, offers little room for those members of lay publics who unequipped to participate or who may actively resist such a positioning. Nor does it acknowledge that open data initiatives are the sites of significant struggle where activist groups contest commercial efforts to co-opt and exploit them (Bates, 2012). The opportunity for lay people and grassroots organisations to make use of the data is limited unless they have technical data science expertise. However it is not only lack of technical expertise that prevents lay people from using big data for their own purposes. Despite much discussion about 'open access' to 'open data' (Davies & Bawa, 2012), such access has lessened as the commercial value of big data has increased. Internet empires (Google, Facebook, Amazon, Twitter and Apple), along with lesser digital corporations and government agencies have control over digital data archives; access to these has become an issue of corporate power and governmental largesse (Fuchs, 2014).

It is therefore difficult for lay people to challenge big data analytics, especially where these are processed algorithmically (Lyon, 2002). As Andrejevic (2014) puts it, there is a 'big data divide', resulting in some actors having ready access to big data and others excluded. Over and above this, as we observed above, lay people can emit digital data or be configured as data subjects without knowing it, let alone consenting to it. Even where people self-consciously digitally monitor themselves, what is done with their data (for example, how it is combined with other datasets) might well lack transparency.

Having noted this, we can also point to members' of the public broader concerns about big data. There is still little research that has investigated what the public makes of big data, aside from reports from privacy organisations and government bodies. What these reveal is deep ambivalence and occasional suspicion (Madden, 2014). Lay people recognize the value of big data for such public goods as maintaining national security, controlling crime, promoting public health, improving healthcare and so on. However many people realize that their data doubles have become commercially valuable and are hostile to the idea that government agencies should sell big data for profit rather than use it for the public good. (The Wellcome

Trust, 2013). The public outcry created by the British government's attempt to capitalize on data from the National Health Service (care.data) is one example (Anonymous, 2014).

In light of the above discussion, we can see that publics can take on numerous epistemic guises in relation to big data. Tentatively we can propose that they are variously 'lay users' (who are users or consumers of big data), 'lay contributors' (who generate and contribute data), 'lay experts' (who are 'digitally literate' and engage with the processes, properties and politics of big data,) and 'expert laypersons' (who actually write code and design technologies). Needless to say, these categories are neither stable nor mutually exclusive, but hopefully they provide a sense of the complexity of what it means to be a 'public' embroiled in the dynamics of big data. Indeed part of the challenge for researchers who are interested in public engagements and entanglements with big data is identifying the ontologies of these categories and their intersections. For example, a 'lay user' of big data is almost inevitably also simultaneously a 'lay contributor': when users browse the internet to search for big data information, their browsing history is recorded and noted by the search engine they use, feeding into a new big data set. 'Big data's understanding of publics' and 'public understanding of big data', in other words, are mutual and co-developing configurations.

Toward a Manifesto for Public Understanding of Big Data

In light of all this, enquiries into how publics engage with, configure, respond to and use big data require new ways of thinking, not least those evidenced in fields such as digital sociology, critical analyses of data, digital cultures, software studies and internet studies. Further, given the ways in which digitized knowledges on people's private habits and preferences are generated and shared as part of big datasets, researching PUBD will need to address contemporary concepts of property, privacy and information security that have previously not much featured in PUS and PEST. Research into public understanding of big data, therefore, involves an engagement with new modes of knowledge production and circulation, new academic literatures and new ways of thinking about data and data practices. That noted, in what follows, we set out the beginnings of a 'manifesto' that addresses what this 'programme of research' might look like. We heuristically organize this in three parts – the Empirical, the Conceptual and the Methodological.

The Empirical

The basic empirical challenge a PUBD faces is how to grasp the sheer variety of data that are being derived, and the multiplicity of means by which these big data are constituted. Leaving aside grand theorizing (Lyotard, 1984; Holton, 1992), the study of the relation between publics and science and technology has tended either to focus on the extent to which publics possess a generic scientific literacy, however that is framed (for example, Durant, 1993), or engaged with specific controversies. PUBD however must address both the many sorts of data that are being 'produced' and their complex and evolving combination. Within science this combination might be said to be grounded in common epistemological principles, as well as complexly mediated through social and institutional configurations (as the literature on interdisciplinarity makes clear – for example Barry et al, 2008). However, big data combination seems to be altogether more opportunistic, piecemeal, mobile and multiple. If

there are 'principles' at play here, what are they? Can they be said to be 'algorithmic' and if so what does this mean? Who produces such 'principles', who implements and circulates them?

The point is that, even at the simplest level, pursuing PUBD requires that we address the issue of what it means to ask members of the public even to think about big data in terms of their *empirical siting and status*. What counts as big data? How do they emerge? Where are they being produced? By whom (eg commercial organizations, state institutions) and by what (eg everyday technologies such as ATMs and surveillance cameras, the emerging internet of things)?

However, insofar as people are also 'prosumers', another terrain of engagement with big data is opened up, namely the study of the practical processes by which a lay person 'emits' digital data, 'receives' data (or its fruits), uses data (in the process emitting more data), and, perhaps on occasion, resists the self-emission and use of data. Here, there are key issues around the 'levels of awareness' that attach to the processes of data emission and use. Many of these data-deriving processes are clearly integral to the socotechnical systems with which we are routinely embroiled (surveillance, web-browsing, credit card use, medical visits, etc, etc), and yet they are also the sometime subject of critical commentary (see above). As such, how we are analytically to address 'awareness' and its relations to understanding and engagement becomes critical. For instance, we might imagine a sort of 'oscillatory awareness' of big data as aspects of its prosumption come into and out of focus, individually and collectively. Thus, issues about privacy, ownership and data exploitation can shift between, or co-exist as - matters of overt concern and matters of routinized utility. How might we empirically access this oscillatory awareness? To the extent that this notion holds water, what does it imply for existing and contested empirical perspectives on the 'public'?

Perhaps more fundamentally, any empirical approach to PUBD must acknowledge that it too is likely to contribute to the making of big data. On this score, PUBD comprises an extreme form of the performativity of social science method (see Law, 2004). To be sure, all such methods heterogeneously construct their 'objects of study'. Indeed, the 'results' derived through such methods - including those of PUS and PEST - can, via various means such as 'governmentality' initiatives (for example Dean, 1999), eventually impact upon publics of various sorts. In the case of PUBD, the very online announcement of a study or the presentation of results can, almost instantaneously, become part of the 'object of study', say, through a platform like Twitter. It is almost as if PUBD becomes its own object of analysis, and, if that is not vertiginous enough, so too does the present discussion of this! Here, it would seem, then that PUBD pointedly encapsulates the processuality of social science research: like 'its' publics it is informant, information and informed in rapid cycles that approach simultaneity.

We can summarise the preceding account as a series of 'manifesto' points:

- Publics are shifting combinations of users and producers
- Empirical data of PUBD is multiple and shifting, iterative and involuted, and can include PUBD itself.

- Expertise is shifting and distributed, spanning the lay and the professional, the human and the nonhuman).
- Big data imply knowledge that is complexly associated with issues of ownership, control and privacy.
- Understanding of, and engagement with, big data needs to be situated in relation to notions of 'awareness' and the various levels at which this operates.

The Conceptual

We have ended up at a point where big data, 'understanding', and publics all seems to be rather complex. This is indicated in the terms we have drawn upon, terms which include: opportunistic, piecemeal, oscillatory, prosumer, mobile, liquid, lively, cooked, rotted, combinatorial, responsive, complex, evolving, vertiginous, processuality. To reiterate, data combined into big data can derive from domains that seem to be rather distant. We can readily envision combining data sets on credit worthiness and visits to medical practitioners to 'deduce' a 'measure' of stress. More exotically, we might imagine the *experimental* combination of data sets, say on household energy consumption patterns and mobile phone use. This is an 'involutionary' combination (for example, Ansell Pearson, 1999) that brings together otherwise disparate data in ways that *potentially* might have value.

The broader point is that under such dynamic and 'involutionary' conditions, such concepts as 'understanding' or 'engagement' are difficult to apply, let alone to operationalize, when they are themselves recursively affected by their object (that is, big data). It seems to us, that of equal importance are such notions as 'affect' and 'imagination'. Caught up in its vertiginous and involutionary processuality, we can envisage the challenges lay people face in grasping – cognitively and socially – big data. In some ways we are affective and affected nodes in the trajectories of data, shaping and shaped by their mobility and combination. To get at this is as much a matter of imagination – how do people 'imagine' the operations of big data? Conceptually, what notions can we draw upon to theorize how publics' affects are mediated through, and reflected by, big data?

As a corollary to the above, big data are also profoundly involved in 'endowing' persons and their sociotechnical relations with particular qualities and identities. Big data, in their specificity around, for instance, security, credit, insurance and health promotion enact publics in terms of their rights to privacy, their capacities to control their finances, their responsibilities to their own and others health, and so on and so forth. It is not exactly difficult to derive an array of case studies or profiles – caricatures even – that congeal to structure, or at least shape, our grasp of some of the operations of big data: the financially irresponsible couple; the environmentally wasteful family; the medically careless individual. Empirically, we might ask members of the public to imagine the sorts of cases/caricatures that have emerged through the operations of big data.

However, taking the foregoing discussion in the round, it would seem that we are faced with patternings of amorphousness and specificity. To borrow from Deleuze and Guattari (1988), we might phrase these patternings in terms of the interplay between deterritorialization

(disordering) and reterritorialization (re-ordering). As they put it: ‘How could movements of deterritorialization and processes of reterritorialization not be relative, always connected, caught up in one another? ...The important point is that... (these)...are not two opposed models...’ (Deleuze and Guattari, 1988, p.20). In the case of big data, we might posit the following patterning: new identities become possible (territorialized) because of the deterritorialization that combines data sets in novel ways, only for those identities to be disputed (deterritorialized) by various groups leading to subsequent efforts to re-territorialize (bring more data sets into novel alignment). It follows that we need to develop concepts which access both the ways in which people are diffusely and specifically affected by big data, and how they are at once rendered ‘emergent’ (that is, deterritorialized) and ‘determinate’ (that is, reterritorialized).

On the basis of this and previous discussions, we can draw out the following manifesto points:

- A PUBD needs to develop concepts capable of addressing the impact of big data, not least insofar as big data entail the combination of more or less disparate data sets yielding more or less stable and novel public identities.
- A PUBD needs to expand its conceptual armoury to address how these big data-derived public identities intersect with existing public identities (eg through health insurance requirements, through social media, etc).
- Given the complexity of public practices in relation to big data, PUBD will need to revisit some of the key ideas of PUS/PEST: how do notions such as public, identity, trust, engagement, folk knowledge, participation and so on hold up in the context of big data?

The Methodological

In light of the various issues discussed above, how a PUBD methodologically addresses big data’s implications for, and impacts on, publics becomes of central importance. As we have seen, the embroilments of ‘lay people’ (and, as we know, this term will require considerable nuancing) with big data (which can include the digital data of a PUBD research event itself) are hugely complex. We are faced with an empirical field marked by: an enormous diversity of data sets that enter into big data; multiple roles that ‘publics’ can adopt or assimilate in relation to big data; the heterogeneity and distributedness of expertise; the muddling of engagement and understanding with diverse or oscillating levels of awareness; and the often opaque associations between the epistemic (knowledge) and the commercial (property, ownership, control). In this context, research events in PUBD are likely to be ‘performative’ (a research method will shape the sort of social data that are produced) and ‘recursive’ (a PUBD study can become part of big data via participants’ use of big data, and thus part of its own object of study) .

Putting this more broadly, we can imagine research in terms of a research ‘event’ in which multiple, heterogeneous elements (minimally participants, research tools, researchers) are brought together to generate ‘data’. On a performative or processual view (Law, 2004), these elements can mutually change – co-become with one another so that the ‘research event’

emerges as something rather different to the ‘research event’ as initially envisaged by the researchers. Under such circumstances (and these apply especially acutely to PUBD), research cannot be a simple matter of *representing* people’s ‘understanding of’ or ‘engagement with’ big data as this is likely to be mobile, multiple, iterative, and emergent. Rather, the empirical task also becomes one of creatively enabling lay people’s imaginative and affective relations to big data to *unfold* (Michael, 2012).

To translate the foregoing into manifesto form, we present the following suggestions:

- PUBD needs to adopt or adapt methodologies that that can encompass the shifting variabilities of publics, expertise, big data and the relations amongst these.
- In particular, such methodologies will need to enable not only the emergence of understandings of, and engagements with, big data, but also facilitate variable forms of awareness of, and imaginative and affective relations to, big data.
- Such methodologies will have to incorporate the peculiar processuality and recursiveness of PUBD research events and address the issue of their own particular forms of performativity in the generation of social data.

While we have suggested some broad programmatic manifesto points for a PUBD methodology, we would also like to provide examples of one particular form that such a methodology might take. To this this end, we provide a short annexe describing some techniques that we developed in an initial pilot study for a wider PUBD programme of research (see Annexe 1).

Concluding Remarks

In this paper, we have attempted to sketch an approach to the Public Understanding of Big Data – one that takes into account the complexities of big data. Obviously, this is an exploratory effort and we are keenly aware that we have barely scratched the surface. Here we present a brief re-consideration of the implications of our discussions of PUBD for PUS/PEST more generally. Firstly, PUBD reminds us that PUS/PEST are parts of wider sociomaterial processes: we cannot do PUS/PEST research without acknowledging that PUS/PEST feeds into a digital knowledge economy, and, as a contributor to big data, impacts on its own object of study. Secondly, the fluid, lively, combinatorial and unfolding character of big data suggests that PUS/PEST might benefit from further developing its approaches to people’s engagement with the future prospects of science and technology. Insofar as technoscientific prospects are digitally mediated they are recursively and complexly part of the knowledge economy. Under these circumstances, lay expectations are likely to be marked by mobility, vagueness, and by what we have called, ‘oscillations of awareness’. Thirdly, and relatedly, this suggests that the imaginative and affective dimensions of PUS/PEST need greater specification. Lay people’s relations to science and technology are, of course, cognitive, social, political, cultural. But as our account of PUBD suggests, they are also affective and imaginative in a knowledge economy marked by de- and re-territorialization, recursion and complexity.

Annex 1

To access the complex dynamic relations between big data, expertise and publics, we drew on the method of ‘cultural probes’ (henceforth, probes). These are techniques developed in the field of design which, as Boehner et al (2012) suggest, are aimed less at deriving accurate interpretations of people’s ‘views’, so much as probing their uncertainty, imagination, ambivalence and so on. More concretely, probes are objects that entail what appear to be nonsensical or strange tasks. Examples include a disposable camera to be used by volunteer to take a photograph of something ‘odd’ like of the spiritual centre of their home, or a ‘Listening Glass’ through which to listen, and on which to record, the sounds of one’s house. Recently probes have also been deployed in ‘engagement workshops’ (Gaver et al, 2015).

It seems to us that something like a ‘probe workshop’ might be a viable means of accessing the affective and imaginative in people’s relations to big data, while also providing insight into people’s understanding and engagement. Here, we propose three probe tasks (developed as part of an ongoing research project). We present these not as definitive examples of the ‘public probing of big data’ as it were, but as initial attempts to develop techniques that meet our Manifesto for a PUBD.

1. *Daily Big Data Task*. This is a collaborative and timed task in which participants work together to fill out a 24 hour time-line of a ‘typical person’s day’ on a paper scroll. They are asked to discuss together and then write on the time-line the kinds of information that is being collected on a ‘typical person’ throughout a day’s activities, including who is collecting the information and what it is used for. This information can include digitized data but also any other form of information. We provide cards suggesting some places where information might be gathered (eg home, workplace, hospitals, supermarkets, parks) and make available post-it notes to be filled and added to the timeline. The rationale is to derive material on how people perceive and imagine the occasions through which data about them is gathered, but also how digital data links to a wider ‘communicative ecology’ (Horst and Miller, 2006). By making this a collective exercise, we aim to enable participants to query the idea of a ‘typical person’s day’.

2. *Digital Profile Card Game*. This is a ‘game’ in which small groups of participants together choose a card from each four stacks. Each stack contains a particular class of information about individuals: demographic, employment, cultural consumption, and health. There is an additional stack of blank cards to which participants can add their own data (eg financial status, travel habits, criminal record). An emerging profile might be a single female aged between 30-40 years, works part-time in a shop, listens to Queen, has diabetes, and often travels to Thailand. We ask participants to discuss how this profile might have been digitally derived, whether it comprises a ‘coherent’ profile and what ‘coherent’ means, and the uses to which such a profile can be put. The rationale here is to enable participants to explore how particular data bases might be combined, how they are generative of identities, and the problematic issues that might arise from these.

3. *Personal Data Machine*. In this exercise, paired participants think about the sort of data they themselves would like collected - either about themselves or about someone else. For

instance, someone might like to find out about their own daily moments of daydreaming; and they might like data collected on their friends' patterns of buying rounds of drinks. We encourage participants to be as innovative or surreal as they wish. This task comes in two parts. In the first, participants take 5 minutes to draw or describe a design to collect digital data about themselves and then reflect on such questions as: What data does your personal data machine collect? What would you do with this data? How might others— family, colleagues, friends, professionals – respond to these digital data machines? For the second task, participants design for digital data collection on others and again reflect on its implications. The rationale here is to encourage participants to imagine the sorts of *potential* that attaches to the derivation and application of big data. Ideally this exercise opens up a space where participants can imagine and affectively respond to some of the extraordinary and unexpected byways that big data might take.

References

- Andrejevic, M. (2013). *Infoglut: How Too Much Information is Changing the Way We Think and Know*. New York: Routledge.
- Andrejevic, M. (2014). The big data divide. *International Journal of Communication*, 8. Retrieved from <http://ijoc.org/index.php/ijoc/article/view/2161>
- Anonymous. (2014). Power to the people (editorial). *Nature*, 505(261). Retrieved from http://www.nature.com/news/power-to-the-people-1.14505?WT.ec_id=NATURE-20140116
- Ansell Pearson, K. (1999). *Germinal Life*. London: Routledge.
- Bates, J. (2012). 'This is what modern deregulation looks like': co-option and contestation in the shaping of the UK's Open Government Data initiative. *The Journal of Community Informatics*, 8(2). Retrieved from <http://ci-journal.net/index.php/ciej/article/view/845/916>
- Baym, N. (2013). Data not seen: the uses and shortcomings of social media metrics. *First Monday*, 18(10). Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/4873/3752>
- Beer, D. (2013). *Popular Culture and New Media: the Politics of Circulation*. Houndmills: Palgrave Macmillan.
- Beer, D., & Burrows, R. (2013). Popular culture, digital archives and the new social life of data. *Theory, Culture & Society*, 30(4), 47-71.
- Best, K. (2010). Living in the control society: surveillance, users and digital screen technologies. *International Journal of Cultural Studies*, 13(1), 5-24.
- Big Data & Society. (2014). About the journal. *Big Data & Society*. Retrieved from <http://bigdatasoc.blogspot.com.au/p/big-data-and-society.html>
- Boehner, K., Gaver, W. and Boucher, A. (2012). Probes. In C. Lury and N. Wakeford (Eds), *Inventive Methods: The Happening of the Social* (pp.185-201). London: Routledge.
- Boellstorff, T. (2013). Making big data, in theory. *First Monday*, 18(10). Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/4869/3750>
- boyd, d., & Crawford, K. (2012). Critical questions for Big Data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, 15(5), 662-679.
- Bruns, A. (2008). *Blogs, Wikipedia, Second Life and Beyond: From Production to Producership*. New York: Peter Lang.
- Callon, M., Lascoumbes, P. and Barthe, Y. (2001). *Acting in an Uncertain World: An Essay on Technical Democracy*. Cambridge, Mass.: The MIT Press.
- Chilvers, J. (2008). Deliberating competence: Theoretical and practitioner perspectives on effective participatory appraisal practice. *Science Technology & Human Values*, 33(2), 155-185.
- Cooper, M. and Waldby, C. (2014). *Clinical Labor: Tissue Donors and Research Subjects in the Global Bioeconomy*. Durham: Duke University Press.
- Davies, T., & Bawa, Z. A. (2012). The promises and perils of open government data (OGD). *The Journal of Community Informatics*, (8), 2. Retrieved from <http://ci-journal.net/index.php/ciej/article/view/929/955>
- Dean, M. (1999) *Governmentality: power and rule in modern society*. London: Sage
- Deleuze, G. and Guattari, F. (1988) *A Thousand Plateaus*. London: Athlone Press
- Fuchs, C. (2014). *Social Media: A Critical Introduction*. London: Sage.

- Durant, J.R. (1993). What is scientific literacy? In J.R. Durant and J. Gregory (Eds), *Science and Culture in Europe* (pp. 129-137). London: Science Museum.
- Gabrys, J. (2014). Programming environments: environmentality and citizen sensing in the smart city. *Environment and Planning D: Society and Space*, 32(1), 30-48.
- Gaver, W., M. Michael, T. Kerridge, A. Wilkie, A. Boucher, L. Ovalle, and M. Plummer-Fernandez (2015). Energy Babble: Mixing Environmentally-Oriented Internet Content to Engage Community Groups. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2015*.
- Gitelman, L. (Ed.). (2013). 'Raw Data' is an Oxymoron. Cambridge, Massachusetts: The MIT Press.
- Hagendijk, R. and Irwin, A. (2006). Public deliberation and governance: Engaging with Science and technology in contemporary Europe. *Minerva*, 44 (2), 167-184.
- Hinchliffe, S., Levidow, L. and Oreszczyn, S. (2014). Engaging cooperative research. *Environment and Planning A*, 46(9), 2080-2094.
- Holton, G. (1992). How to think about the 'anti-science' phenomenon. *Public Understanding of Science*, 1, 103-128.
- Horst, H.A. and Miller, D. (2006). *The Cell Phone: An Anthropology of Communication*. Oxford: Berg.
- Irwin, A. and Michael, M. (2003). *Science, Social Theory and Public Knowledge*. Maidenhead, Berks.: Open University Press/McGraw-Hill.
- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Kitchin, R., & Lauriault, T. (2014). *Towards critical data studies: charting and unpacking data assemblages and their work*. Geoweb and Big Data. Book chapter in press. University of Nebraska Press. Nebraska.
- Lash, S. (2006). Life (Vitalism). *Theory, Culture & Society*, 23(2-3), 323-329.
- Law, J. (2004). *After Method: Mess in Social Science Research*. London: Routledge.
- Lupton, D. (in press). *The Quantified Self: A Sociology of Self-tracking Cultures*. Cambridge: Polity Press.
- Lyon, D. (2002). Everyday surveillance: Personal data and social classifications. *Information, Communication & Society*, 5(2), 242-257.
- Lyon, D., & Bauman, Z. (2013). *Liquid Surveillance: A Conversation*. Oxford: Wiley.
- Lyotard, J-F. (1984). *The postmodern condition: A report on knowledge*.
- Madden, M. (2014). Public perceptions of privacy and security in the post-Snowden era. *Pew Research Internet Project*. Retrieved from <http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/#>
- Marres, N. (2012). *Material Participation: Technology, the Environment and Everyday Publics*. Basingstoke: Palgrave.
- Michael, M. (2009). Publics performing publics: Of PiGs, PiPs and Politics. *Public Understanding of Science*, 18, 617-631.
- Michael, M. (2012). "What are we busy doing?": Engaging the Idiot. *Science, Technology and Human Values*, 37 (5), 528-554.
- Nowotny, H., Scott, P. & Gibbons, M. (2001) *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity Press.

- Räsänen, M., & Nyce, J. M. (2013). The raw is cooked: data in intelligence practice. *Science, Technology & Human Values*, 38(5), 655-677.
- Ritzer, G. (2014). Prosumption: evolution, revolution, or eternal return of the same? *Journal of Consumer Culture*, 14(1), 3-24.
- Rogers, R. (2013). *Digital Methods*. Cambridge, MA: The MIT Press.
- Ruppert, E. (2011). Population objects: interpassive subjects. *Sociology*, 45(2), 218-233.
- Thrift, N. (2014). The 'sentient' city and what it may portend. *Big Data & Society*, 1(1). Retrieved from <http://bds.sagepub.com/content/1/1/2053951714532241.full.pdf+html>
- Totaro, P., & Ninno, D. (2014). The concept of algorithm as an interpretative key of modern rationality. *Theory, Culture & Society*, 31(4), 29-49.
- The Wellcome Trust (2013). *Summary Report of Qualitative Research into Public Attitudes to Personal Data and Linking Personal Data*. London: The Wellcome Trust.
- Welsh, I. and Wynne, B. (2013). Science, scientism and imaginaries of publics in the UK: Passive objects, incipient y threats. *Science as Culture*, 22 (4), 540-566.
- Wynne, B.E. (1995). The public understanding of science. In S. Jasanoff, G.E. Markle, J.C. Peterson and T. Pinch (Eds). *Handbook of Science and Technology Studies* (pp. 361-388). Thousand Oaks, Cal.: Sage.