**Preface**

We thank the European Association for the Study of Science and Technology (EASST) for hosting the conference track that resulted in this book and for providing a congenial and inspiring collegial environment at the conference overall. We also thank all participants in the track for their contributions as presenters, discussants, and debaters. Thanks, too, to those who contributed their work to this volume, especially those who also did internal review work on chapters. Last but not least, thanks to our publishers and their anonymous reviewers whose comments have helped us immensely in improving and completing the text.

**Chapter 1**

**The politics of Big Data: Big Data, Big Brother?**

**Introduction to the volume**

Ann Rudinow Sætnan (0000-0003-4147-7836), Ingrid Schneider (0000-0002-1270-1942) and Nicola Green ( )

**Our focus and goal**

Kranzberg’s 'First Law of Technology’ teaches us that ‘[t]echnology is neither good nor bad; nor is it neutral’ (Kranzberg 1986: 547). By this, he means that technologies engage with complex social ecologies to produce social results, but that these results are not entirely controlled by the technologies' human interactors[[1]](#endnote-1). Having learned also from Barad (2007), we could put this that technologies *intra-act* with human societies – both taking shape and shaping in each moment of entanglement. These mutual shaping moments occur whether we reflect on them or not; we can enter into them more or less watchfully, less or more carelessly. Furthermore, the mutual shaping happens continuously, throughout a technology’s social history. As long as we engage with a technology, we continue to shape it and be shaped by it. By engaging reflexively in that reshaping process, we can also reshape the process itself.

Big Data is a case in point. Big Data is constituted by a nexus of technologies – data processing hardware and software and a myriad of digitised apparatuses and processes networked through a constant data flow, each element embedded into the varied ecologies of current society. This means we are constantly shaping and being shaped by Big Data. While this basic claim can be made of any technology, from the stone wheel to the atom bomb, it seems today that Big Data is taking form as a particularly powerful reshaping force. In this reshaping process, benefits and harms are unevenly distributed across social groups and across different aspects of social life. Forms and degrees of influence over the shaping processes also vary and are unevenly distributed. The more we learn about those distributions – both of benefits and harms being produced through the implementation of Big Data, and of in what ways and how effectively individuals, groups and governments have so far been engaged in shaping Big Data – the more effectively we can engage in the shaping process and its outcomes in future.

This book focuses on the shaping of and by Big Data, approaching that theme from many angles. Here in the introduction we will first describe what it is the contributors share in our conceptualisation of that theme. We will then describe a key element of what our various angles of approach have in common as well as how they differ in that regard. Finally, we will give readers a brief presentation of the structure of and contributions to the volume.

Our contributions to this book first came into conversation with one another when we all responded to a call for papers to form a conference track under the open and questioning title ‘Big Data, Big Brother?’ As those conversations proceeded, we saw that, however varied our contributions were in some regards, they shared a sub-theme: they all addressed issues regarding the policies and politics of Big Data. In other words, the very variety of our approaches served to tighten our shared focus on the political aspects of Big Data. At the same time, that variety of approaches demonstrated the breadth of the political focus, namely the breadth of what can be labelled as ‘politics’ and ‘policies’. Accordingly, in this introduction to the volume, we will discuss two concepts that define its scope:

1. What we mean by ‘Big Data’, and
2. What we mean by ‘politics and policies’.

**Big Data – does it exist, and if so, what is it?**

In the discussion after the final session of our conference track, one member of the audience challenged us on the very concept of Big Data: ‘Isn’t Big Data all just hype?’ he asked, ‘Does it even actually exist?’[[2]](#endnote-2) Our answer is that it both is and isn’t ‘all just hype’, Big Data does and doesn’t really exist; or rather, it certainly exists, but one might well ask in what sense and to what extent it exists.

At the very least, Big Data exists as a socio-technical imaginary that serves as a meta-narrative to capture the present and future of digitisation, datafication, and globalised networks. For instance, boyd and Crawford (2012: 663) define Big Data as ‘a cultural, technological, and scholarly phenomenon’ that rests on the interplay of technology, analysis, and mythology, referring to the latter as the ‘widespread *belief* that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the *aura of* [emphases added] truth, objectivity, and accuracy.’ When that imaginary is expressed, for instance in the form of a definition, Big Data also takes on a semantic existence. From there, either as an imaginary or as a semantic reality, Big Data exists in a phenomenological sense, i.e. as subjective perceptions that form an interpretive framework and a basis for action. And finally, Big Data exists because (as many of the empirical chapters will show) we can see those imaginaries and frameworks being acted upon. When Big Data imaginaries, semantics and interpretive frameworks are enacted and the actants (both human and non-human) of those frameworks engaged, Big Data takes on material form(s). Yet these forms do not always fully realise the shapes and traits imagined for them.

So what are these imaginaries and frameworks? A clear definition would make the phenomenon semantically more real while also bounding it in against conceptual drift, but definitions and undefined usages vary widely. It is perhaps unfortunate that there is no apparent consensus on a definition, however many definitions do seem to refer to some form of observable materiality.

Some definitions, as we shall see below, focus on size. Although hyped as revolutionary in that regard (See e.g. Mayer-Schönberger and Cukier’s (2013) *Big data: A revolution that will transform how we live, work, and think*), in a sense there is nothing new about Big Data. There is a long history of dealing with amounts of information that at some point seemed too vast to handle (see for instance Hacking 1982). Cynically speaking, the only thing ‘new’ about Big Data is the catchy name. Less cynically, what’s new is its scale. Stipulating some minimum size that would deserve the name ‘Big’ might at first glance seem to set a precise boundary vis à vis older data practices. For instance, Intel defines Big Data organisations as those ‘generating a median of 300 terabytes (TB) of data weekly’ (Intel 2012). And, certainly there are such organisations.

However, size is a moving target. As data capacities rise and costs fall, 300 TB/week may soon seem not so big after all. Another dimension is relative size. Rather than a fixed number, the current Wikipedia (2017) article defines Big Data as ‘data sets so large or complex that traditional data processing application software is inadequate to deal with them’.[[3]](#endnote-3) Again, yes, certainly there are organisations that struggle to make sense of their data due to the sheer size of the databases they deal with.

For a company such as Intel that sells data capacity, size – be it a fixed figure or a relative target – may be a useful definition criterion. For us as philosophers/historians/social scientists of technology, size, to be relevant, should preferably (at least also) relate to meanings, practices, and the social and ethical consequences thereof. Laney (2001) points to three purported traits of Big Data that may relate to meanings, practices and consequences. While not presented as a definition of Big Data, the three traits – volume, velocity and variety, also known as ‘the three Vs’ – have been much used as one (e.g. by UK POST 2014 and by Mayer-Schönberger and Cukier 2013).

For instance, it is used – and expanded upon – by one of the few book-size publications providing a critical analysis of Big Data. Based on literature reviews, Rob Kitchin (2014: 1-2) defines Big Data as:

- huge in volume, consisting of terabytes or petabytes of data;

- high in velocity, being created in or near real-time;

- diverse in variety, being structured and unstructured in nature;

- exhaustive in scope, striving to capture entire populations or systems (n=all);

- fine-grained in resolution and uniquely indexical in identification;

- relational in nature, containing common fields that enable the conjoining of different data sets;

- flexible, holding the traits of extensionality (can add new fields easily) and scalability (can expand in size rapidly)

Discussing these traits one by one, Kitchin points out how each is not necessarily some sort of data-based ‘superpower’, but also problematic. For instance, volume challenges data storage and analysis capacity, although less so as capacities grow. Velocity challenges analysts’ ability to define and capture relevant moments in ever-changing data flows and trends. Variety challenges the ability to curate – categorise, standardise, stabilise – data for collation, comparison, and analysis. The existence of whole conferences on the curation challenges of Big Data databases (for example IDCC 2017) points to the existence of such issues.

Some, e.g. IBM (undated) add a fourth V for Veracity, but this is a misnomer. There is no evidence that massed data from multiple sources are ‘truer’ than data collected in traditional scientific endeavours. In fact, in their presentation of the fourth V, IBM subtitles ‘veracity’ as ‘uncertainty of data’. Thus, veracity has to be questioned as data may be inaccurate or unreliable.

Other analysts have added Variability as a fifth V: Data qualities and results may change, for example over time, and newly-collected data or novel ways of using existing data, for instance by applying new analysis methods, or by linking together data sets to give a different picture. Variability can also be addressed by asking previously un-asked questions, or by correlating data formerly unrelated (UK POST 2014).

Google has defined Big Data self-referentially by word-clouding search terms that co-occur with it (Ward and Barker 2013). That they managed to perform such an analysis is doubly self-referential evidence that the phenomenon exists, as a search term and as an approach to repurposing and analysing masses of stored data, in this case data from Google searches. We might even say the results were triply self-referential, as they pointed to other phenomena enmeshed with the concept and practices of Big Data: According to Ward and Baker (2013), Google users who seek information about Big Data often also seek information about Big Data analytics, in general or with specific reference to database technologies such as the data storage and retrieval architecture style known as NoSQL or data storage and analysis software programs such as Hadoop.

Big Data analytics, rather than database size alone, is also where the Big Data hype now focuses. This shift in hype focus may be similar to the focus shift from genomics to proteomics (see for example Fujimura 2005, Webster 2005). One way of interpreting such shifts is as Latour (1987: 114-121) describes the rhetorical mobilisation and hiding of barriers (moving attention from one problem to the next, and thus away from unachieved ‘final’ goals) as a strategy to maintain member loyalty in socio-technical networks. In the case of Big Data we could paraphrase these as follows: ‘You now have access to vast amounts of data and are only more confused? Don’t abandon the bandwagon! We only need to develop a little more self-learning software; then we will reach your goals.’ However, as we shall see, the problems may have stayed on the bandwagon too, especially if they were not merely about size, but also about analysis, ethics, legality, actionability and so on.

Continuing from how big, via how difficult and how to do, many influential definitions (often implicit rather than explicit and, ironically, often anecdotal or casuistic rather than statistics-based) focus on what Big Data analytics have achieved. For instance, Mayer-Schönberger and Cukier (2013) define Big Data as ‘The ability of society to harness information in novel ways to produce useful insights or goods and services of significant value’ (ibid.: 2) and ‘…things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value’ (ibid.: 6). Here we are getting to definitions we might challenge as pointing to a mythical object rather than a material one, to ‘air ware’ and hollow promises (or threats), or to boyd and Crawford’s (2012: 663) aforementioned definition of Big Data as a belief system.

Although Mayer-Schönberger and Cukier (2013) temper their Big Data success tale with the caution that privacy may suffer, their emphasis is nevertheless on Big Data’s purported advantages, summarising them as the very definition of Big Data. Others, e.g. Anderson (2008), mention no disadvantages whatsoever. While some (e.g. boyd and Crawford 2012; Kitchin 2014; Sætnan in this volume) go further in their critique of Big Data, examining even its success claims, most authors are less critical. And, judging by numbers of citations (according to both Google Scholar and ISI Web of Science), Big Data enthusiast texts clearly have greater traction than critiques. Through that traction, Big Data – however mythical its results may be – exists as a political, societal and commercial force, a force resulting in attempts to *perform* Big Data by (re-)collecting and attempting to (re-)use vast amounts of data.

We can readily find such arguments pitched towards businesses. Just do an internet search for ‘Big Data solutions’ and there they are. One company, called Big Data Solutions, advertises that,

We support our customers in capitalizing the data and combine it with other information in order to produce improved business insight. […] Whether you are currently struggling with dirty, unorganized or unknown data assets looking to uncover opportunities and leveraging big data or embarking on your first master data management initiative, we can help you unlock the true value of your enterprise data. (Big Data Solutions undated)

Another company, SAP, advises potential client firms to, ‘Discover the Big Data solutions that give you 360-degree insight into your business and help you find hidden growth opportunities within the digital noise’ (SAP undated). And so the ads read, from company after company offering Big Data hardware, software, cloud storage and/or analytics services. Across all these ads, the message is clear: since you have all these data anyway, why not crunch their numbers and squeeze more profits from them? To put it even more succinctly – the data are yours; use them. From this, and playing as fast and loose with spelling as the age-old concept of ‘Three R’s’ for Reading, wRiting, and aRrithmetic, we could add yet another V to the description of Big Data: aVailability. A word of caution, however: this V does not necessarily mean that Big Data capitalises on data that are freely available. Data available to a given Big Data operator *may* be freely and openly available, but much is privately held, proprietary, confidential, even classified as secret. This of course entails that data now have commercial value, so Big Data operators may also be crunching numbers on data they have purchased. And, as with anything of commercial value, data are also tempting objects for crime: they may be hacked, stolen, forged, falsified … and even when legally held, they may be illegally sold, shared, repurposed and/or combined (see Sætnan in this volume). So, aVailability highlights simply that Big Data capitalises on whatever data an operator has access to.

Summing up, we can now tentatively[[4]](#endnote-4) define Big Data as *the collection and aggregation of large masses of (publicly, commercially, proprietarily, and/or illicitly) available data and its analysis[[5]](#endnote-5), largely in the form of correlation, pattern-recognition, and predictive analysis*. And by this definition again, yes, Big Data definitely exists. It exists in various sizes – some bigger, some smaller in terms of numbers of data points collected and numbers of persons covered by the data set, some simply stored unanalysed, some analysed through sophisticated algorithms including ‘self-learning’ AI programs. The data are collected by individuals, organisations, firms, states. Analysis is performed for commercial, medical, research, political, policiary, military, and criminal purposes. Sometimes the sole new purpose of a collection by one party is to sell the data on to yet other parties.

Much of this data is ‘personal data’[[6]](#endnote-6). According to the 1995 EU Data Protection Directive 95/46/EC (European Parliament 1995), ‘personal data’ is:

Any information relating to an identified or identifiable natural person (‘Data Subject’); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity*.* (95/46/EC article 2, §a)

This definition is similarly incorporated into the various national Data Protection Acts. For instance the Irish Data Protection Act of 1988, section 1, was amended in 2003 to state that:

‘personal data’ means data relating to a living individual who is or can be identified either from the data or from the data in conjunction with other information that is in, or is likely to come into, the possession of the data controller; (Irish Statute Book)

The phrases ‘or indirectly’ and ‘likely to come into the possession of’ are key here. Perhaps at a given moment only a few factors that might identify you, the data subject, can be found within a given database, let’s say your age and the purchases you made at the pharmacy yesterday. Suppose these two factors are in the database, but not your gender, your address, your hair colour, your name … nothing else that could identify you, and there are many people your age who were pharmacy customers that day, so your anonymity is safe. So far. But then, suppose it would be a simple matter for the controller of that database to find your address, or gender, or place of employment and link those together with your age and any other data recorded about you in the database. They might still not have your name, but they would have enough factors to render all the information about you in that database (including that which was previously considered safely anonymised) personally identifiable.

This is made even more explicit in the EU General Data Protection Regulation (GDPR 2016), which takes force in May 2018. Here an ‘identifiable person’ is defined not only by that person being linked to a name or number, but is any natural person ‘who can be identified, directly or indirectly, in particular by reference to *an identifier* [emphasis added]’ with identifiers being no so much defined as exemplified through the list ‘such as a name, an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person’ (GDPR 2016, Article 4(1)).

Looking from these definitions of personal data back at Kitchin’s (2014) definition of Big Data, and noting his points that such databases are ‘fine-grained’, ‘indexical’ and ‘relational in nature’, we can see how even the most innocuous and anonymous bits of data emanating from our daily activities quickly become personal, even intimately invasive, when linked together in Big Data collections. This concerns not only data we in some sense ourselves report. In the Internet of Things, sensors and smart devices may provide data and metadata that – even though not originally personalised – may be used for profiling and for re-identification of individuals and for classifying and predicting their behaviour. So even just our handling of daily objects can contribute to the masses of data that can potentially be traced back to us.

Historically, the most important large data sets (such as censuses or tax records) have been about persons as citizens, and have thereby contributed to constituting them (us) as such. Currently, and increasingly, Big Data will derive from our interactions with technical artefacts. The availability of sensors and our cultural fascination with things ‘high-tech’ and ‘smart’ may encourage manufacturers and platform providers to monitor use of their products by clients in ways that could create commercial value. Such sources and uses of data have hardly any historical precedent, and need to be scrutinised. For instance, how do they constitute us as users, as customers, as clients, as integrated aspects of technologies, etc. – and how might they change us as citizens?

So far we have concluded that yes, Big Data does exist. However, the caution remains that Big Data also in some sense does *not* exist. If we define Big Data according to what it is sometimes hyped to be – that some shadowy ‘They’, through the vast collection and incredibly accurate analysis of secondary data, now know everything about everybody – then no. Big Data is not a collection and accurate analysis of everything about everybody, as will be discussed by Sætnan and by Matzner in this volume. Some might say ‘not yet’, but we would rather refrain from making alarmist, hyped-up predictions. We also believe that there can never be a collection of everything about everybody. Some persons, for at least some of their data, will always elude collection, and some aspects of our persons are not collectable in data form. And yet, dangerously, ‘practically everything about everybody’ does seem to be what some users and abusers, regulators and anti-regulatory activists, Big Data recorders and recorded believe it to be (e.g. Whitaker 1999; Cole 2014) – and is certainly what some Big Data promoters would have them believe (e.g. Clarke 2013). Furthermore, some collections do come closer to ‘everything about everybody’ than many feel comfortable with: for instance the NSA’s data collections as revealed by Edgar Snowden (Greenwald 2014; Harding 2014), or Facebook’s collection of not only our Facebook posts, searches, and ‘likes’, but also much of what we do on-line even when not logged on to Facebook (Gibbs 2015), or Google’s collection of our entire Google search histories (Galperin 2012). Confronted with such vast and detailed collections of data, and with claimed capabilities and explicit aims of analysing these data, it is hardly surprising that social scientists would want to study Big Data. But what should we be asking and how should we go about seeking answers?

**Politics as a perspective on Big Data**

The researchers who presented at the ‘Big Data, Big Brother?’ track, most of whom have contributed to this volume, represent a variety of social science and humanities fields and disciplines: Anthropology, Cultural Studies, Gender Studies, Informatics, Law, Media Studies, Philosophy, Political Science, Science and Technology Studies, Sociology. When defining disciplines, there is a focus on how they differ in their core concerns. Here, we found that we shared an understanding of what we were (and still are) studying and why it was (still is) important to study it. We found that we were primarily studying Big Data in terms of how it serves to (re-)distribute power in social relationships. This means we shared a concern for the *politics* of Big Data, for instance as politics are defined in the Collins English Dictionary (2017):

Politics (pɒlɪtɪks) 1. plural noun. Politics are the actions or activities concerned with achieving and using power in a country or society.

According to that definition, we share a general concern with politics. We deepen that perspective when, paper by paper, our respective political concerns focus on specific, more narrowly defined, aspects of the broad political field. Some focus on governmental policies seeking to promote, use, or regulate Big Data. These fit with the concept of politics according to the topmost – presumably most common – definitions in the Merriam-Webster Dictionary (2017):

1 a**:** the art or science of government

b**:** the art or science concerned with guiding or influencing governmental policy

c**:** the art or science concerned with winning and holding control over a government

2**:** political actions, practices, or policies

The Merriam-Webster definition of politics overlaps with what some would separate out as ‘policies’. The divide between these two is contested, and since the contest devolves into one of a semantic convention, only language history can settle the question, and only the end of humanity can finalise language history. For now, however, one fairly conventional way of separating between these concepts is to think of ‘politics’ as the matter of principles and priorities, the debate or controversy side of the coin, and of ‘policy’ as the resulting pragmatic efforts of governing bodies to regulate and order society. Some use the two terms the other way around. There are, of course, many more ways of conceptualising the divide. For instance, one might think of politics as the negotiative process of overcoming conflicts and policies as a way to intervene and even exacerbate conflicts.

Politics, whether viewed as debate and controversy or as negotiation, can be studied from many angles. Political party programs offer a view of their principles and goals, policy documents often begin with statements of overarching aims, policies can be analysed to reveal the conflicting interests negotiated to achieve them, media texts can be a window into as well as a driving force for public opinion – just to mention a few such perspectives. All of these are represented in this volume. In Chapter 4, Strauß analyses the pressures of various interests that have led to securitisation being a dominant theme in governance, including Big Data policies. In Chapter 6, Rieder examines EU policy documents, focusing on their visionary side. In Chapter 10, Tøndel and Sætnan follow public discourse on Big Data and surveillance through the window of mass media texts.

The policy side of the politics/policy coin can itself include at least two coin sides when it comes to Big Data. One side is efforts to regulate Big Data, as discussed by several chapters in this volume, most thoroughly by Schneider, who takes us through national and supra-national regulatory efforts and legal sanctions in Chapter 8. The other side is governments’ own usage(s) of Big Data as an administrative tool, on which see Ochs (Chapter 13 in this volume), and also Yeung (2017).

Of course, all the chapters in the volume address concerns of politics and/or policies in one or more of these senses of the terms. Politics and policies as matters concerning governments and electoral processes is the central theme of Section 2: Big Data Policies: Politics of Governance and Regulation, where it is approached from a number of angles, ranging from overarching political imaginaries to undergirding public opinions. Policies are most directly discussed by Pasquale in Chapter 7, where he discusses regulators’ struggles to keep up with the speed, spread and technological evolution of data-driven media when it comes to imposing editorial controls and citizen protections; and, by Schneider in Chapter 8, where she gives a detailed account of Big Data regulations and controls at both national and supra-national levels and as practiced by various branches of government. Both politics and policies also come up in the other sections. In Chapter 4, Strauß discusses why securitisation has such dominance among political goals and why Big Data is so often taken as a ‘silver bullet’ policy tool to achieve national security. In Chapter 12, Fleischhack gives an inside view of certain NGOs’ attempts at ‘taming’ Big Data; then in Chapter 13, Ochs discusses why this ‘self-protection’ approach is both propounded and subverted by many governmental agencies, in part because Big Data is a tool governments themselves use in public administration. In Chapter 14, Noorman et al. discuss government policies aimed at promoting open access to research data, then follow up by discussing practical and ethical issues raised from Academe’s ‘grass roots’ that have slowed the implementation of such policies. In other words, the book, examines multiple branches and levels of government in terms of their Big Data politics, policies and practices, be those aimed at promoting, using, and/or regulating Big Data.

Governmental agencies are not the only collective environment where technologies engage with social ecologies with results that affect social distributions. Humans act collectively in contexts outside of governments and electoral politics as well. Businesses and markets are one example, and several chapters deal with how businesses affect and are affected by Big Data. In fact, most chapters touch on this aspect, but some contain a specific focus on it, as when Ingrid Schneider (Chapter 8) discusses network effects in digital platforms and the forces of oligopolisation.

Another channel for collective human action (including intra-action with Big Data) is non-governmental organisations (NGOs). In Chapter 12, Fleischhack presents and analyses activities of several groups focused precisely on Big Data issues, in particular on individual citizens’ data privacy. The NGO’s ‘digital literacy’ and self-protection approach is not, of course, the only strategy with regard to privacy, a theme touched on by nearly all chapters in the volume.

But, much as governmental agencies are not the only channels for ‘achieving and using power in a country or society’ (i.e. not the only channels for politics as defined by Collins English Dictionary), neither are corporations or NGOs, however influential, the only alternative channels. Public discourse does not distribute power solely through its effects on electoral politics or legislative priorities. Discourse is a productive source of power in its own right through its regulation of what statements may be made and by whom regarding a given discourse topic (Foucault 1980). While it is primarily other concepts from Foucault that are used by authors in this volume, it is in keeping with this concept of discourse – as something that tentatively imposes an order and thus a worldview regarding what is shown, by whom, and by what means – that Big Data discourses are discussed in several chapters in the volume. Thus, discourses affect the mutual shaping of Big Data and society – not only through their influence on electoral outcomes or official policies, but also in terms of how they delineate the subjects and objects of Big Data, setting an agenda for what its issues are and who speaks authoritatively to those issues.

Of course, politics is not only about public office or public discourse. Politics, even when accessed through public office or public discourse, is in the final analysis about power distributions in our everyday lives and relationships. The second-wave feminist slogan ‘the personal is political’ emphasised this fact on two counts: firstly that power differentials in everyday lives and personal relationships are important, and secondly that the responsibility for resolving them should not be relegated to individuals alone but is a legitimate matter for political organisation (Hanisch 2006). Thus, the everyday practices of Big Data systems – including not only how Big Data practices impact on ordinary, grassroots citizens, but also how they are enacted and/or resisted individually and collectively – is an aspect of Big Data politics. We focus on this aspect in several chapters dealing with the everyday practices of Big Data, especially in Section 3: Performance is Political: Big Data Practices, Performance, and Resistance. Aspects of everyday life discussed in this section include: how surveillance of our video watching preferences are recursively used to script videos we watch (Chapter 11), the online practices and competencies of children (Chapter 12), and the work practices and constraints of academics as these affect participation (or not) in open data schemes.

We also take heed that technologies are not merely passive mediators of their human designers’ intentions. In his seminal article,[[7]](#endnote-7) ‘Do artefacts have politics?’ (Winner 1980), Langdon Winner points out how redistribution of social goods and influence is sometimes designed into technologies intentionally. While such intentions may be inscribed into them, ascribed to them, de-inscribed from them and so on (Akrich 1992, Akrich and Latour 1992), social distribution outcomes are not a result of design alone. For one thing, designers’ scripts will encounter ‘user scripts’ (Gjøen and Hård 2002, see also Oudshoorn and Pinch 2005) once a technology is implemented[[8]](#endnote-8). Furthermore, technologies’ and other non-human actants’ materiality may itself enable or resist humans’ intentions (see e.g. Callon 1986). Looking at this from another angle: designers, users, technologies and the objects they describe ‘intra-act’ with one another, all emerging re-shaped from that intra-action (Barad 2007). Thus, the book also discusses technological aspects of Big Data, how these sometimes resist fulfilling Big Data designers’ intentions, and how systems and citizens re-shape one another when we are subjected to Big Data scrutiny. That too is an aspect of Big Data politics, an aspect dealt with especially in Section 1: Principles and Paradigms: Questioning the tenets of Big Data, while the user side of the equation is highlighted most in Section 3: Performance is Political - Big Data Practices, Performance, and Resistance. Intra-actions of users (a video streaming service), end users (video viewers), and non-human actants (Big Data and screenplay scripts) are the focus of Chapter 11, ‘No (big) data no fiction?’ by Rocco Bellanova and Gloria Gonzalez Fuster.

Of course, any volume must find space between its covers. Therefore, there are inevitably topics within the theme of Big Data politics that we have covered only cursorily. One of these is how Big Data – or any data, large or small – influence not only data politics, but also politics overall. The theme is not entirely absent from the book, as the role of statistics in the practices of governance is touched on by Ochs in Chapter 13. For more thorough coverage of this theme, we can recommend a number of sources: See, for instance Hacking (1990) and Desrosières (1998) on the history of statistics, Rose (1991) on the relation between democracy and practices of calculation, Barry (2002) about the ‘political’ and ‘anti-political’ use (and reappropriation) of quantification systems, Sætnan, Lomell and Hammer (2011) about intra-actions between statistics, policies, and practices, Bellanova (2017) about European data protection and the possibility to open a space of political critique with regard to data-driven governance, as well as Yeung (2017) mentioned above.

Finally, we regret that this volume on politics and policies of Big Data, a focus arising out of a shared interest in how Big Data practices re-shape power relations, does not include chapters dedicated specifically to the analysis of Big Data in intra-action with major intersectional categories such as gender, race, or sexual orientation. These categories too refer to what are essentially power-distributive relations. Such categories are acknowledged in several chapters in the volume. We would have wished, however, to have found and included analyses focused specifically on intersectionalities. That we were unable to do so may reflect a gap in Big Data research as a whole, although we do know of research on intersectional aspects of surveillance more generally – e.g. Browne (2015) on surveillance and race, van der Meulen and Heynen (2016) on surveillance and gender, and Lyon (2003) on surveillance as social sorting. Hopefully, our own ‘missing’ papers will soon be published elsewhere, extending the discourse on the politics of Big Data beyond the bindings of this book. But for now, let us return to what the book does contain:

**The volume in brief**

As has already been outlined, the main body of the volume is divided into three sections. Section 1 begins with a critical analysis of some of the key principles and epistemics of Big Data – i.e. what knowledge is Big Data meant to generate and by what means? In Chapter 2, ‘The haystack fallacy, or why Big Data provides little security’, Sætnan evaluates the claims of central proponents of Big Data that the data in their sheer vastness and variety constitute nearly everything knowable, and that statistical principles developed for data samples are therefore obsolete. Sætnan also discusses problematic consequences of margins of error when Big Data are deployed into security contexts. Next, in ‘Grasping the ethics and politics of algorithms’, Matzner presents a critical analysis of the next step in Big Data – algorithmic analytics – and finds that the problems remain. In chapters 4 and 5 – ‘Big Data – within the tides of securitisation?’ (Strauß) and ‘Surveillance as critical paradigm for Big Data?’ (Matzner) – we see the political forces driving the expansion and deployment of Big Data into national security measures, in spite of all the dangers such deployment carries with it. Matzner, however, also interrogates to what extent ‘surveillance’ as a critical paradigm, often expressed in the metaphors of the panopticon or the omniscient eye of Big Brother, is adequate to capture the challenges of Big Data. While critical of security applications of Big Data and of its potential for social and political harms, all four chapters point to potential benefits of Big Data, for instance as a market tool and for generating scientific hypotheses.

In Section 2, the emphasis is on public policies more broadly and discourses potentially affecting these. In chapter 6 (‘Tracing Big Data imaginaries through public policy: The case of the European Commission’), Rieder analyses official documents of the European Commission to show what sociotechnical imaginaries underlie the Union’s policy goals for a ‘Big Data future’. Pasquale (Chapter 7, ‘The automated public sphere’) examines more specifically policies relating to the algorithmic automation of the public sphere, and subsequent effects on opinion-building and political decision-making. Pasquale points out that consumer protection and media regulatory authorities must intervene and that new methods of monitoring and regulating Big Data will be needed in order to keep up with technological and economic developments. Schneider (Chapter 8, ‘Bringing the state back in: Big Data-based capitalism, disruption, and novel regulatory approaches in Europe’) explains the mechanisms behind the rise of digital platforms and the impacts of algorithmic ranking, rating and scoring on social sorting and the democratic social order. She also provides readers with a detailed, comprehensive, and critically analytical overview of EU and national regulatory policies relating to the political economy of Big Data. Schneider’s overview demonstrates that many regulatory policies have been aimed at the protection of fundamental rights, including privacy. In Chapter 9, ‘Rear window – transparent citizens versus political participation’, Simões and Jerónimo argue that privacy is a mediating factor; the more important issue is citizens’ autonomy, with loss of autonomy in the wake of Big Data posing a serious threat to democracy. Finally in this section, Tøndel and Sætnan (Chapter 10, ‘Fading dots, disappearing lines – Surveillance and Big Data in news media after the Snowden revelations’) examine the extent to which Edward Snowden achieved his stated goals for leaking NSA documents and the former editor-in-chief of *The Guardian*, Alan Rusbridger, his goals for agreeing to publish these highly critical revelations about bulk data surveillance. Did the Snowden leaks change public discourse on Big Data and surveillance, and if so, did that discourse lead to changes in public policy?

With a nod to Hanisch (2006), ‘The Personal is Political’, we have given Section 3 the title ‘Performance is Political: Big Data Practices, Performance, and Resistance’. The four chapters in this section focus on three examples of everyday Big Data practices in different spheres of activity. In chapter 11, ‘No (big) data no fiction? Thinking surveillance with/against Netflix’, Bellanova and Gonzalez Fuster give an in-depth analysis of how the content streaming platform has used Big Data from its customers/viewers to tailor the content of a series on Big Data surveillance in politics. Netflix thus creates a recursive loop of watched 🡪 watching 🡪 watchers 🡪 debating watching and, potentially, of surveillance opinions 🡪 shaping surveillance presentations 🡪 shaping surveillance opinions. In Chapter 12, ‘“Data trainings” in German schools – Learning empowerment from hackers’, Fleischhack reports from participant observation of grassroots activist workshops aimed at motivating and enabling users of social media and other digital technologies to better protect their own and others’ privacy. Fleischhack points to both resources for and limitations of such individual-based protections. In the next chapter (Chapter 13, ‘Self-protection beyond the self: Collective privacy practices in (Big) datascapes’), Ochs further analyses how individualized ‘informational self-protection’ strategies have come to dominate governmental policies, and how their logic is seriously flawed since self-protective tools and encryption practices are of necessity collective in nature. Turning to another inherently collective social endeavour, namely science, Noorman, Wessels, Sveinsdottír and Wyatt (Chapter 14, ‘Understanding the “open” in making research data open: Policy rhetoric and research practice’) examine the technological, organisational and moral dimensions of the arguments for, arguments against, and pragmatic problems of sharing research data. Of course, Big Data is deployed into and/or has repercussions within far more spheres of daily activity than these. Some, such as commercial applications, have been much discussed elsewhere. Others remain ‘dark continents’ on our virtual maps of the state of knowledge.

We could have no reasonable expectation of covering the whole terrain of Big Data practices. Our expectation is that this mix of fields and experiences – some eclectic and some almost universally relevant across applications – can inspire further reading and research. With this in mind, Green has written our concluding chapter as a methodological postscript: ‘Big Data’s methodological challenges’. The title could alternatively have been ‘Big Data’s reflexive methodological inspirations’, or at least the reflexive inspirations from our own findings regarding Big Data. For instance, if we accept Ochs’s findings (and we do accept them) that privacy protection must inherently be a collective enterprise, and reflect that the same is true of research, then we must ask ourselves how we might go about building a community effort to research Big Data. Might it involve learning from unexpected actors, such as the hackers in Fleischhack’s chapter? How would we (as Noorman et al. point out that we must) make our data shareable within Big Data research as a community project? Green points to problems of in/visibilities, scale, location, and approach – problems that are perhaps universal in science, and that Big Data can help us resolve if we hold it up not only as an object of curiosity, but also as a research(ing) subject and a mirror.

**References:**

Akrich, M (1992). The De-Scription of Technical Objects. In Bijker, W.E. and Law, J. (eds.) *Shaping Technology/Building Society*. Cambridge, MA: MIT Press: 205-224.

Akrich, M. and Latour, B. (1992) A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies. In Bijker, W.E. and Law, J. (eds.) *Shaping Technology/Building Society*. Cambridge, MA: MIT Press: 259-264.

Anderson, C. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired Magazine*: 16.07. [online] Available at: <http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory>. [Accessed 01 Oct 2015].

Barad, K. (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durham & London: Duke University Press.

Barry, A. (2002). The anti-political economy. *Economy and Society,* 31**(**2): 268-284.

Bellanova, R. (2017). Digital, politics, and algorithms: Governing digital data through the lens of data protection. *European Journal of Social Theory,* 20**(**3): 329-347

Big Data Solutions (undated). Big Data Solutions [online] Available at: <http://www.bigdatasolutions.fi/>. [Accessed 31 Oct 2017].

boyd, d. and Crawford K. (2012). Critical Questions for Big Data, *Information, Communication & Society*, 15:5, 662-679, page 663)

Browne, S. (2015). *Dark Matters: On the Surveillance of Blackness*. Duke University Press.

Callon, M. (1986). Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. In Law, J. (ed.) Power, action and belief: a new sociology of knowledge? London, Routledge, 1986, pp.196-223. [online] Available at: [http://www.vub.ac.be/SOCO/tesa/RENCOM/Callon%20(1986)%20Some%20elements%20of%20a%20sociology%20of%20translation.pdf](http://www.vub.ac.be/SOCO/tesa/RENCOM/Callon%20%281986%29%20Some%20elements%20of%20a%20sociology%20of%20translation.pdf) [Accessed 13 Oct 2017].

Clarke, R.Y. (2013). Smart Cities and the Internet of Everything: The Foundation for Delivering Next-Generation Citizen Services. White paper sponsored by Cisco. *IDC Government Insights*. [online] Available at: <http://119.15.167.84:8080/share/proxy/alfresco-noauth/api/internal/shared/node/q9Ij_C2XQhS0ElSMm-jJnA/content/GI243955.pdf>. [Accessed 01 Oct 2015].

Cole D. (2014). ‘We kill people based on metadata’. *The New York Review of Books*, 10 May 2014. [online] Available at: <http://www.nybooks.com/blogs/nyrblog/2014/may/10/we-kill-people-based-metadata/>. [Accessed 22 Sep 2015].

Collins English Dictionary (2017). Politics. [online] Available at: <https://www.collinsdictionary.com/dictionary/english/politics>. [Accessed 31 Oct 2017].

Desrosières, A. (1998) *The Politics of Large Numbers. A History of Statistical Reasoning*. Cambridge, MA: Harvard University Press.

European Parliament (1995). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. [online] <http://ec.europa.eu/justice/policies/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf>. [Accessed 25 October 2017].

European Parliament (2016). **Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available at** <http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf> **(Accessed 25 October 2017).**

Feenberg, A. (1999). *Questioning Technology*. London: Routledge.

Feenberg, A. (2010). Ten paradoxes of Technology. *Techné*, 14(1): 3-15. Originally presented as a keynote lecture in 2009. [online] Available at: <https://www.youtube.com/watch?v=-HzJ_Jkqa2Q>. [Accessed 23 Oct 2017]

Foucault, M. (1980). Power/Knowledge: Selected interviews and other writings, 1972-1977. New York: Pantheon Books. [online] Available at: <https://monoskop.org/images/5/5d/Foucault_Michel_Power_Knowledge_Selected_Interviews_and_Other_Writings_1972-1977.pdf>. [Accessed 25 October 2017].

Fujimura, J.H. (2005). Postgenomic futures: translations across the machine-nature border in systems biology. *New Genetics and Society*, 24(2): 195-226.

Galperin, E. (2012). How to Remove Your Google Search History Before Google's New Privacy Policy Takes Effect. Entry on webpages of Electronic Frontier Foundation. [online] Available at: <https://www.eff.org/deeplinks/2012/02/how-remove-your-google-search-history-googles-new-privacy-policy-takes-effect>. [Accessed 24 Sept 2015].

Gibbs, S. (2015). Facebook 'tracks all visitors, breaching EU law'. People without Facebook accounts, logged out users, and EU users who have explicitly opted out of tracking are all being tracked, report says. [online] Available at: <http://www.theguardian.com/technology/2015/mar/31/facebook-tracks-all-visitors-breaching-eu-law-report?CMP=share_btn_tw>. [Accessed 31 Oct 2017].

Gjøen, H. and Hård, M. (2002). Cultural Politics in Action: Developing User Scripts in Relation to the Electric Vehicle. Science, Technology & Human Values, 27(2): 262-281.

Greenwald, G. (2014). *No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State*. Metropolitan Books;

Hacking, I. (1982). Biopower and the avalanche of printed numbers. *Humanities in Society,* 5(3-4): 279-295.

Hacking, I (1990). *The taming of chance*. Cambridge: Cambridge University Press.

Hanisch, C. (2006). The Personal Is Political. The Women’s Liberation Movement classic with a new explanatory introduction. [online] Available at: <http://www.carolhanisch.org/CHwritings/PIP.html>. [Accessed 11 Oct 2017].

Harding, L. (2014). *The Snowden Files: The Inside Story of the World's Most Wanted Man*, Vintage Books.

IBM (undated). The 4 Vs of big data. [online] Available at: <http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data.jpg>. [Accessed 22 Sep 2015].

IDCC (2017) International Digital Curation Conference (IDCC), Digital Curation Centre. [online] Available at: <http://www.dcc.ac.uk/events/internationa-digital-curation-conference-idcc>. [Accessed 31 Oct 2017].

Intel (2012). Big Data Analytics. Intel’s IT Manager Survey on How Organizations Are Using Big Data. [online] <http://www.intel.com/content/dam/www/public/us/en/documents/reports/data-insights-peer-research-report.pdf>. [Accessed 31 July 2015].

Irish Statute Book (2003). Data Protection (Amendment) Act 2003, Section 2. [online] Available at: <http://www.irishstatutebook.ie/eli/2003/act/6/section/2/enacted/en/html#sec2>. [Accessed 02 Oct 2015].

Joerges, B. (1999). Do Politics Have Artefacts? *Social Studies of Science,* 29**(**3): 411-431.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). [online] Available at: <http://journals.sagepub.com/doi/abs/10.1177/2053951714528481>. [Accessed 25 Oct 2017].

Kranzberg, M. (1986). Technology and History: “Kranzberg’s Laws”. Technology and Culture, 27(3): 554-560.

Laney, D (2001). 3D Data Management: Controlling Data Volume, Velocity and Variety. [online] Available at: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. [Accessed 31 July 2015].

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge: Harvard University Press.

Lomell H.M., Dahl, J.Y. and Sætnan, A.R. (2007). Å se og bli sett: Kommunikative aspekter ved videoovervåkning” [To see and be seen: Communicative aspects of video surveillance], in Spilker & Levold (eds.) *Kommunkasjonssamfunnet. Moral, praksis og digital teknologi*. [Communication Society: Morals, Practices and Digital Technologies.] Oslo: Universitetsforlaget, 105-119.

Lyon, D. (2003). *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination.* London and New York: Routledge.

Mayer-Schönberger V. and Cukier K. (2013). *Big data: A revolution that will transform how we live, work, and think*. London: John Murray Publishers, Ltd.

Merriam-Webster (2017). Definition of Politics. [online] Available at: <https://www.merriam-webster.com/dictionary/politics>. [Accessed 31 Oct. 2017].

Oudshoorn, N. & Pinch, T. (eds.) (2005). *How Users Matter: The Co-Construction of Users and Technology,* Cambridge, MA: MIT Press.

Rose, N. (1991). Governing by numbers: figuring out democracy. *Accounting, Organizations and Society,* 16**(**7): 673-692

Saetnan, A.R., Lomell, H.M. and Hammer, S. (eds.) (2011). *The Mutual Construction of Statistics and Society*. London: Routledge.

SAP (undated]. Big Data Software Solutions SAP. [online] Available at: <https://www.sap.com/trends/big-data.html>. [Accessed 31 Oct 2017].

UK POST (2014). Big Data: An Overview. UK Houses of Parliament, Parliamentary Office of Science and Technology (POST). July 2014.

van der Meulen, E. and Heynen, R. (2016). *Expanding the Gaze: Gender and the Politics of Surveillance*. Toronto: University of Toronto Press.

Ward J. S. and Barker A. (2013). Undefined by data: A survey of big data definitions. [online] Available at: <http://arxiv.org/abs/1309.5821>. [Accessed 22 Sept 2015].

Webster, A. (2005). Social science and a post-genomic future: alternative readings of genomic agency. *New Genetics and Society*, 24(2): 227-238.

Whitaker, R. (1999). *The End of Privacy: How Total Surveillance is Becoming a Reality*. New York: The New Press.

Wikipedia (2017). Big data. [online] Available at: <https://en.wikipedia.org/wiki/Big_data>. [Accessed 31 Oct 2017].

Winner, L. (1980). Do Artifacts Have Politics? *Daedalus* 109(1): 121-136.

Yeung, K. (2017). ‘Hypernudge’: Big Data as a mode of regulation by design. *Information, Communication & Society,* 20**(**1): 118-136.

1. Marvin Kranzberg, an historian, drew these ‘laws’ – which he says are not laws in a legal or moral sense so much as a ‘series of truisms’ (Kranzberg 1986: 544) – from a long career’s experience researching technologies. The six laws are: 1) Technology is not good, or bad; nor is it neutral. 2) Invention is the mother of necessity. 3) Technology comes in packages, big and small. 4) Although technology may be a prime element in many public issues, nontechnical factors take precedence in technology-policy decisions. 5) All history is relevant, but the history of technology is most relevant 6) Technology is a very human activity – and so is the history of technology (Kranzberg 1986). Similar points are made by the philosopher Andrew Feenberg, for instance in his book *Questioning Technology* (1999) as well as in the article ‘Ten paradoxes of Technology’ (2010). [↑](#endnote-ref-1)
2. Not an actual quote, just our synopsis of his point. [↑](#endnote-ref-2)
3. One could also consider the issue of database size relative to data handling purpose, as when data protection authorities deny a data handling concession on the basis that the prospective data handler is requesting permission to gather and store more data than their stated purpose warrants. Data sets regarded as overlarge in such a context, however, are not used as defining cases for Big Data. Rather, promoters of Big Data tend to ignore this issue altogether, focusing on the opportunities large data sets provide rather than the on the legality of vast data collections. [↑](#endnote-ref-3)
4. Authors may use other definitions in their respective chapters, but for now this one will serve to project an image of our object of study. [↑](#endnote-ref-4)
5. This is often secondary analysis (i.e. analysis for new purposes, other than those for which the data were originally produced and collected). [↑](#endnote-ref-5)
6. There are exceptions, such as data generated by sensors monitoring natural or atmospheric phenomena like the weather or pollution, or technical aspects of manufacturing processes. [↑](#endnote-ref-6)
7. The article sparked considerable debate and the theme of politics as an inherent aspect of technologies continues to this day. See, for instance, Joerges (1999). [↑](#endnote-ref-7)
8. Or, for that matter, non-user scripts, as resistance to implementation is also of consequence. [↑](#endnote-ref-8)