Original Research Article

BIG DATA & SOCIETY

Big Data & Society
July–December 2014: 1–13

DOI: 10.1177/2053951714541861
bds.sagepub.com
SAGE

Surveillance, Snowden, and Big Data: Capacities, consequences, critique

David Lyon

Abstract

The Snowden revelations about National Security Agency surveillance, starting in 2013, along with the ambiguous complicity of internet companies and the international controversies that followed provide a perfect segue into contemporary conundrums of surveillance and Big Data. Attention has shifted from late C20th information technologies and networks to a C21st focus on data, currently crystallized in "Big Data." Big Data intensifies certain surveillance trends associated with information technology and networks, and is thus implicated in fresh but fluid configurations. This is considered in three main ways: One, the capacities of Big Data (including metadata) intensify surveillance by expanding interconnected datasets and analytical tools. Existing dynamics of influence, risk-management, and control increase their speed and scope through new techniques, especially predictive analytics. Two, while Big Data appears to be about size, qualitative change in surveillance practices is also perceptible, accenting consequences. Important trends persist – the control motif, faith in technology, public-private synergies, and user-involvement – but the future-orientation increasingly severs surveillance from history and memory and the quest for pattern-discovery is used to justify unprecedented access to data. Three, the ethical turn becomes more urgent as a mode of critique. Modernity's predilection for certain definitions of privacy betrays the subjects of surveillance who, so far from conforming to the abstract, disembodied image of both computing and legal practices, are engaged and embodied users-in-relation whose activities both fuel and foreclose surveillance.

Keywords

Surveillance, privacy, Big Data, control, ethics, Snowden

Introduction: Snowden disclosures and Big Data

The Snowden revelations about National Security Agency (NSA) surveillance, starting in June 2013, along with the ambiguous complicity of internet companies and the international controversies that followed illustrate perfectly the ways that Big Data has a supportive relationship with surveillance. Words such as "bulk data" and "dragnet" and "mass surveillance" more than hint that processes referred to as "Big Data" are in play, producing expanded and intensified surveillance. The rapid and widespread adoption of what are called Big Data practices signal profound changes for individuals, for the dynamics of both public and private sector organizations, for the relation of citizen to state, and for society at large.

However, for a fuller understanding of Snowden's revelations and Big Data surveillance several matters have to be unpacked, not least the questions of the *socio-technical* character of Big Data, how several of the Snowden revelations demonstrate dependence on Big Data techniques and which have a highly significant impact for understanding the character of surveillance today. Of course, some of what Snowden has revealed involves targeting but the main focus here is on Big Data techniques. Beyond this, it is vital to consider what is meant by the controversial key concepts,

Queen's University, Ontario, Canada

Corresponding author:
David Lyon, Queen's University, University Avenue, Kingston, ON, Canada K7L 3N6.
Email: lyond@queensu.ca

"Big Data and surveillance." What follows is a provocative introduction to some key issues raised by the social and political realities of this conceptual conjunction, prompted by the Snowden revelations.

Big Data, first, may be best thought of as "the capacity to search, aggregate and cross-reference large data sets" (Boyd and Crawford, 2012: 663). There is of course a range of ideas, practices, metaphors, software, and techniques bundled together in those two deceptively straightforward-sounding words. For one thing, Big Data practices occur in a variety of contexts (Meyer-Schönberger and Cukier, 2012) and one big mistake is to imagine that similar kinds of ends and possibilities of success are in view whatever the context. Consumer marketing, health care, urban policing, and anti-terrorism – to take four popular potential and actual application sites for Big Data – are not the same and practices that may in some respects be acceptable in one (say, marketing) may erode rights and deny human dignity in another (say, anti-terrorism) (Mosco, 2014: 177–205). If there are potential benefits or harms, that is, they are not the same in each area.

The second crucial concept is surveillance, that can be understood as any systematic, routine, and focused attention to personal details for a given purpose (such as management, influence, or entitlement; see Lyon, 2007: 13–16). This too is a broad definition that needs some tightening for the present purpose. Our task here is to examine how far Big Data intensifies certain surveillance trends associated with information technologies and networks (see Bennett et al., 2014), and is thus implicated emerging configurations of power and influence. Of course, as political-economic and socio-technological circumstances change, so surveillance also undergoes alteration, sometimes transformation. Classically, studies of surveillance suggest that a shift in emphasis from *discipline* to *control* (Deleuze, 1992; Haggerty and Ericson 2000) has been a key trend associated with the increasing use of networked electronic technologies that permit surveillance of mobile populations rather than only those confined to relatively circumscribed spaces, and depend on aggregating increasingly fragmented data. Surveillance practices have been moving steadily from targeted scrutiny of "populations" and individuals to mass monitoring in search of what Oscar Gandy calls "actionable intelligence" (2012: 125) and Big Data surveillance exemplifies this.

Two main questions are addressed here: One, in what ways and to what extent do the Snowden disclosures indicate that Big Data practices are becoming increasingly important to surveillance? Queries about Big Data practices in relation to surveillance and public concern about the activities of the NSA predate Snowden, of course (Andrejevic and Gates, 2014). But Snowden's revelations have brought them into the public eye as never before. Two, if Big Data is gaining ground in this area, then how far does this indicate changes in the politics and practices of surveillance? Are new trends, or the augmentation of older ones, visible here? We shall explore these questions in respect to the *capacities* of Big Data and their social-political *consequences* before commenting on the kinds of *critique* that may be appropriate for assessing and responding to these developments.

The Snowden disclosures

The first item revealed by Edward Snowden on 6 June 2013 and published in *The Guardian* (UK) was that the NSA, using an order from the Foreign Intelligence Surveillance Court (FISC), had required the telecommunications giant Verizon to hand over metadata from millions of American's phone calls to the Federal Bureau of Investigation and the NSA (Greenwald, 2013). Verizon itself was forbidden to disclose to the public either the order or the request for customer records.

The next day, articles in the *Washington Post* and *The Guardian* detailed how the PRISM program seemed to give the NSA direct access to the servers of some of the biggest technology companies, including Apple, Facebook, Google, Microsoft, Skype, Yahoo, and YouTube. Encryption and privacy controls were circumvented with the help of the companies (Gellman and Poitras, 2013). In the UK, the Tempora program appeared to be even more like a dragnet as it gave similar access to GCHQ (General Communications Headquarters; the UK partner of the NSA in the "Five Eyes"). Together, their cable and network tapping abilities are called "Upstream" and can intercept any internet traffic. The database that allows the information to be extracted in real time is called "XKeyscore" (Lanchester, 2013). The revelations have continued and Snowden himself has said (in early 2014) that some of the most striking disclosures are yet to come.

The surveillance practices revealed by Snowden show clearly if not completely that governments – especially American, British, Canadian, and possibly other agencies – engage in astonishingly large scale monitoring of populations, and also *how* they do it. On the one hand, the NSA engages contractors to share the burden of their work and also gathers and mines user data collected by other corporations, especially telephone, internet, and web companies. And on the other, this kind of surveillance also means that the NSA and similar agencies watch for cookies and log-in information. They thus use data derived from the use of devices such as cell phones or geo-locating social media

sites. What users unknowingly disclose on those platforms – such as Facebook or Twitter – or when using their phones, is usable data for "national security" and policing purposes. But more importantly from a Big Data perspective, metadata (see the discussion below) relating to users is gleaned without their knowledge from the simple use of these machines. There are thus at least three significant actors in this drama, government agencies, private corporations and, albeit unwittingly, ordinary users.

What holds these groups together, in a sense, is the software, the algorithms, the codes that allow users' data to be systematically extracted or disclosed, analyzed, and turned into what the data collectors and others, such as the NSA, hope will be actionable data. In other words, it is the (big) data practices that different kinds of operations have in common. As Snowden himself said in a 10 June 2013 video, the "... NSA targets the communications of everyone ..." then "... filters, analyzes, measures them and stores them for periods of time simply because it's the easiest, most efficient and most valuable way of achieving these ends" (Greenwald et al., 2013). The NSA thus depends on codes, the algorithms, plus the witting or unwitting cooperation of both telephone and internet corporations in order to do surveillance. Individual users may play a part, too, but their role is hardly one of conscious actors in the drama. This already goes beyond what many once imagined was direct and specifically targeted relationships by state agencies of individuals, to mass surveillance, dependent on a close liaison with corporate bodies and on the self-recording devices used in everyday communications and transactions.

The gathering of national intelligence in the U.S. is a mammoth undertaking, worth over US$70 billion per year (FAS, 2014) and involving extensive links with universities, internet companies, social media, and outside contractors – such as Booz Allen Hamilton that employed Edward Snowden and from which Snowden illegally conducted his removal of sensitive data. If nothing else, the economic value of these operations indicates how much emphasis is placed on data processing by government agencies and in turn by global corporations. But what kinds of data are sucked up so voraciously by these organizations with such sophisticated processing power?

The word that has perhaps appeared most in relation to the Snowden revelations is "metadata." This term refers – rather imprecisely – to the "data about data" such as the IP address, the identity of the contact, the location of calls or messages, and the duration of the contact. However, metadata takes many forms, well beyond communications. For example, automatic license plate recognition systems or word-processing programs also generate metadata (Newell, forthcoming). While specific cases of monitoring the content of phone calls and examining text messages exist as well, the extremely large-scale collection and analysis of metadata characterizes many of the disclosures about the kinds of activities with which the NSA is engaged. When the Snowden revelations began in June 2013, governments and agencies were quick to dismiss them by downplaying the significance of metadata.

In the U.S., the collection of metadata was permitted after 9/11 under the "Section 215 Bulk telephony metadata program" but it is unclear how far similar such programs extend to other countries such as Canada or the UK. However, it was revealed in 2014 that the Canada Border Services Agency made 19,000 requests for subscriber data in one year but this and other related Canadian agencies are under no statutory requirement to say how often such requests are made or for how much data (Freeze, 2014). More specifically, a program that featured in the news media as Canada's CSEC (Communications Security Establishment of Canada) collecting data from airport Wifi systems was actually a general means of identifying travel patterns and geographic locations using ID data (that is, metadata) in conjunction with a database of IP addresses supplied by the company Quova over a two-week period in January 2014. What this shows is how data are analyzed, rather than just the fact of its collection (Schneier, 2014). Such data may be used, for instance, to set up an alarm when a "suspect" enters a particular hotel, or to check on someone – a kidnapper, maybe – who may have repeatedly visited a particular location. But it takes little imagination to think of other potential uses for such datasets.

This is why security critic Bruce Schneier cuts through the obfuscations to state unequivocally that "metadata is surveillance."[1] As he also observes, while the mass media accounts focus on *what* surveillance data are being collected, the most significant question is *how* the NSA analyzes those data. On the one hand, the nearly five billion cell phone records collected by the NSA each day by tapping into cables that connect mobile networks globally can reveal personal data about where users are located, anywhere in the world. The NSA can attempt to track individuals to private homes and can also retrace earlier journeys, whenever the phone is on, because phones transmit location data whether or not they are in use. On the other hand, the NSA also analyzes patterns of behavior to reveal more personal information and relationships between different users (Gellman and Soltani, 2013). The latter is more subtle, but in the Big Data world, more significant.

These pattern-seeking processes are the ones where Big Data practices really come into their own.

For example, the NSA program, known as ''Co-Traveler,'' uses highly sophisticated mathematical techniques to map cell phone users' relationships, superimposing them on others to find significant intersections and correlations. Co-Traveler is meant to search for the associates of foreign intelligence targets, although domestic users' data are also garnered ''incidentally'' and the foreign sweeps are so broad that they are bound in include Americans on a mass scale. This is the searching, aggregating, and cross-referencing process referred to above, that characterizes some of the technical aspects of Big Data.

Several key surveillance trends (see Bennett et al., 2014) are augmented by the advent of Big Data. Just two are mentioned here. One is that contemporary surveillance expands exponentially – it renders ordinary everyday lives increasingly transparent to large organizations. The corollary, however, is that organizations engaged in surveillance are increasingly invisible to those whose data are garnered and used. This ''paradox'' is deepened by the advent of Big Data (Richards and King, 2013). A second trend is that the expanding securitization of daily life prompts the use of extended surveillance, from neighborhoods and travel arrangements to large sporting and entertainment events. The quest of ''national security'' breeds Big Data, particularly through efforts to preempt security breaches by a form of anticipatory surveillance described, somewhat vaguely, by the Department of Homeland Security as ''connecting the dots.''

Of course, the surveillance implications of the use of Big Data – such as using metadata – are just one dimension of new ways of structuring information in a digital age. The present task is not to catalogue potentially beneficial aspects of Big Data but rather to focus attention on what sorts of surveillance issues are raised – especially ones that prompt civil liberties or privacy questions – in new ways by this re-structuring of information.

Big Data surveillance

The Big Data/surveillance link was recognized by US President Obama on 17 January 2014, when he called for a ''comprehensive review of Big Data and privacy'' following the Snowden leaks (White House, 2014). It was further acknowledged when the US proposed new rules governing bulk data collection by the NSA of the phone calling habits of Americans (Savage, 2013). The once-secret bulk phone records problem was what had most alarmed privacy advocates when the Snowden leaks began in 2013 and now the president proposed that it should be curtailed, along the lines of a dated European data retention directive. But the media rhetoric surrounding this suggests a fairly conventional understanding of surveillance that does not fully grasp the Big Data aspects of the bulk phone records.

Surveillance constantly undergoes change and is currently being reconfigured in several respects, some of which alter its character. In particular, different kinds of data are now being captured and used in new ways, which prompts some to distinguish between surveillance as targeted practices over against fresh form of dataveillance (van Dijck, 2014). Not only are data captured differently, they are also processed, combined, and analyzed in new ways. Social media that appeared on the scene at roughly the same time as responses to 9/11 boosted the ''surveillance state,'' are now the source of much data, used not only for commercial but also for ''security'' purposes. The buzzword is ''datafication,'' which points to the ways that for many businesses, the information infrastructure is their heart (Bertolucci, 2013). Ordinary users' social activities are sucked up as data, quantified and classified, making possible real-time tracking and monitoring.

It goes beyond this, however. With Big Data practices, for example, personal data – now including identifiable metadata – are not collected for certain limited, specified, and transparent purposes, which are the goals of data protection and privacy advocates. Rather, Big Data reverses prior policing or intelligence activities that would conventionally have targeted suspects or persons of interest and then sought data about them. Now bulk data are obtained and data are aggregated from different sources *before* determining the full range of their actual and potential uses and mobilizing algorithms and analytics not only to understand a past sequence of events but also to predict and intervene *before* behaviors, events, and processes are set in train. Both corporate and government aspects of this raise questions for analysis and critique.

Preemptive approaches in security and policing, that depend on prediction, have been growing steadily since the 1990s and were extensively augmented after 9/11, are a bureaucratic incentive to over-collect data, especially in security and law enforcement. Perhaps even more important to cost-cutting government departments, the falling cost of processing power is a strong inducement to use new data analytics in a number of fields (Bankston and Soltani, 2014). It is not hard to find extravagant promises that real-time data analytics will transform aspects of retail, manufacturing, health care, and public sector organizations. But despite the determined and well-informed activities of data protection and privacy advocates over a number of years and in several countries, any countervailing focus on the contribution Big Data may make to reducing democratic freedoms, reconfiguring privacy and indeed, redefining the role of information in contemporary life, is still muted and marginalized.

Thus stated, the problem is that basic alterations in surveillance and legal expectation are occurring in a context that celebrates rather than carefully assesses Big Data. The differences between Big Data applications are crucial here. For example, Ian Kerr and Jessica Earle (2013) distinguish helpfully between three kinds of prediction: consequential, where one aim is to help clients or users to choose what is likely to be beneficial to them, preferential, illustrated by marketers trying to second guess our desires from our browsing behavior, and preemptive, where is a deliberate intention to reduce someone's range of options. In the context of law and justice, the latter raises fundamental issues of privacy and due process. Where legal systems are based on an after the fact system of penalties or punishments, the turn to one based on future-oriented preventative measures is of huge import, not least for those rendered unable to understand or contribute meaningfully to the process.

This is why the Snowden revelations offer a unique opportunity to grapple with Big Data surveillance in a systematic way. Of course, this phenomenon did not appear overnight, fully formed. It represents the confluence of many streams and is itself better thought of as a fluid form that constantly changes its character than as a relatively solid set of surveillance relations that positions and governs the subject in a disciplinary fashion. Introducing the language of surveillance to Big Data discussions challenges the practices often described in epistemologically naïve and politically disingenuous ways. But equally, examining how Big Data practices are affecting the character of contemporary surveillance obliges students of surveillance to reconsider what is happening, particularly across different surveillance domains.

Capacities

The term ''Big Data'' suggests that size is its key feature. Massive quantities of data about people and their activities are indeed generated by Big Data practices and many corporate and government bodies wish to capitalize on what is understood as the Big Data boom. As with many other single aspects of this phenomenon, however, the idea of size both yields important clues and, on its own, can mislead. While the *capacities* of Big Data practices (including the use of metadata) intensify surveillance by expanding interconnected datasets and analytical tools, this tells only a part of the story.

Drawing on a number of sources, Rob Kitchin argues that Big Data has several crucially important characteristics: huge volume, consisting of terabytes or petabytes of data; high velocity, being created in or near real time; extensive variety, both structured and unstructured; exhaustive in scope, striving to capture entire populations of systems; fine-grained resolution, aiming at maximum detail, while being indexical in identification; relational, with common fields that enable the conjoining of different data-sets; flexible, with traits of extensionality (easily adding new fields) and scalability (the potential to expand rapidly) (Kitchin, 2014: 262).

Data sources may be thought of under three main headings each of which may be applied in surveillance contexts: directed, automated, and volunteered (Kitchin, 2014, forthcoming). In the first, a human operator obtains the data, obvious examples being CCTV systems or police seeking, say, vehicle ownership records. In the second, the data are gathered without a human operator intervening; traces are recorded routinely from transactions with banks or consumer outlets and communications, using cellphones above all. In the third, data are in a weak sense ''volunteered'' by the user who gives out information on social media sites and the like. Of course, social media users do not necessarily think of their activities in terms of volunteering data to third parties (Trottier, 2012) but this is an accurate way of understanding surveillance data gathering in this context.

Clearly, one of the surveillance trends amplified by Big Data practices is the increased integration of government and commercial surveillance. Big Data may also be thought of in terms of its promised economic rewards. As Bruce Schneier observes, the term Big Data could be viewed as placing today's data operations in the same kind of category as ''Big Pharma'' or ''Big Oil'' (Schneier, 2012), where the corporate strategy behind the new practice is the decisively significant factor and when ''big'' refers to the economic worth of the data commodity. This dimension is very important to any analysis, as Viktor Mayer-Schönberger and Kenneth Cukier (2012) show. They argue that the ''Big Data revolution'' is based in part on new data management techniques that permit analysis beyond ''rows and tables'' to dispensing with ''hierarchies and homogeneity'' but also on internet companies collecting vast troves of data and having ''a burning financial incentive to make use of them'' such that they became leading users of the latest processing technologies, sometimes superseding others that had decades more experience (2012: 6).

Understood thus, the capacities of Big Data surveillance take on some new meanings. The enthusiasm for commercial uses of Big Data is shared by those in the security field, thus stimulating further integration of these activities. In a Big Data context, the same data are increasingly used for different purposes. This is more than a change of context that might alter how data subjects might construe their privacy or how

legal limits on secondary use might be stretched. Rather, the same commercial data may be given new meanings in the security realm, combined, and connected in novel ways. Thus the capacities of Big Data may also be seen to allow new forms of inferential reasoning that Louise Amoore calls ''data derivatives'' (2011). This is of critical importance in the security-surveillance context because such associations and links, however trivial and improbable, may be given new meanings that are cut off from the values that once made sense of them and the identifiable subjects whose activities generated them in the first place.

Consequences

As with the capacities of Big Data, given the current volatility of the field it is hard to tell exactly what the consequences of widespread adoption of Big Data will be for surveillance. Important trends will probably persist, including the quest for control through surveillance, an almost naïve faith in technology that inhibits the search for low-tech or no-tech alternatives, public-private synergies that benefit government, corporation – and sometimes citizens – and the involvement of internet users in surveillance processes as ''prosumers.'' At the same time, the reinforced future-orientation is likely to exacerbate the severance of surveillance from history and memory and the assiduous quest for pattern-discovery will justify unprecedented access to data. It is likely that existing dynamics of influence, risk-management, and control will increase their speed and scope through new techniques, especially predictive analytics, but what specific trends will be accentuated as Big Data practices expand?

There are three key ways in which commitment to Big Data practices seem to be shifting the emphasis of surveillance and this is clear in the Snowden disclosures. They are stated here and discussed below. First, given that Big Data involves the amplified use of algorithms for analytics, an increasing reliance on software for surveillance and a concomitant reliance on what might be called a ''human-algorithm'' relationship that shapes the ways that human subjects are treated by surveillance systems. Such *automation* tends to diminish opportunities for discretion within systems. Second, Big Data practices increasingly tilt surveillance operations to focus on the future more than on the present and the past. In the context of neo-liberal governance, this *anticipation* is likely to place more weight on surveillance for managing consequences rather than research on understanding causes of social problems such as crime and disorder. The third area is *adaptation,* the propensity for analytics to be treated as if methods can be transferred successfully and with little risk from one field to another. The enthusiasm for Big Data ''solutions'' may lead to the inappropriate transfer of techniques from one field to another.

Automation

The combination of readily available software and its relatively low price is an incentive to choose technical solutions over more labor-intensive ones in surveillance practices as in other fields (see Bankston and Soltani, 2014 on how this affects police location tracking). This means that automated surveillance will become an increasing possibility. At the same time, greater data storage capacity means that larger and larger amounts of data are collected before their use has been ascertained (Savage and Burrows, 2007), the consequences of which are unknown as yet. What we do know, however, is that who makes decisions about algorithms and datasets will have the capacity to make a difference in these emerging scenarios (Glennon, 2014).

The automation of surveillance must also be seen as an aspect of the way that surveillance occurs as a routine management procedure. Evelyn Ruppert rightly warns against panoptic panics regarding government surveillance that suggest sinister state attempts to keep close watch on all citizens. The automating of surveillance is part of the kinds of cost-cutting and efficiency exercises that have dominated the public administration for decades. So far from there being an ''all-knowing state, what we have instead is a plethora of partial projects and initiatives that are seeking to harness ICTs in the service of better knowing and governing individuals and populations'' (Ruppert, 2012: 118).

And if the process is not panoptic then it is not directly disciplinary either. The ''modulating'' controls described briefly but evocatively by Gilles Deleuze (1992) are more in view here than direct concern with discipline or with the behavior of individuals. Ruppert argues that databases work with an ontology of subjects that creates profiles – data bodies or data doubles – based on their activities, connections, performances, transactions, and movements that relate to government. These data ''make up'' the people in the system purview, in ways that are constantly shifting, fluctuating. In this way, a neo-liberal logic of control fits neatly with the ways that individuals are ''made up'' by data. If the role of ''data doubles'' in determining the life-chances and choices of individuals was a major concern of an earlier phase of surveillance studies (see e.g. Lyon, 2001) then its Big Data magnification will likely intensify such concerns, both analytically and in terms of critique and political contestation. ''Data doubles'' becomes a *double-entendre*.

The kind of ''soft biopower'' (Cheney-Lippold, 2011: 166) associated with Big Data is at work in marketing as in parallel ways to those ''harder'' forms found in

"national security." Marketing moved from demographic (Gandy, 1993) to more psychographic categories in the 1990s and then as marketing went online, was able to use search histories to create further consumer clustering, superimposed on the former categories. Algorithms are used increasingly to target particular kinds of consumers in relation more to real-time web use than to the older categories of census and postcode. This contributes to cybernetic-type control, where what is assumed to be normal and correct behavior is embedded in circuits of consumer (or employment, health, or education) practices. This is also significant for what Snowden has revealed, as we shall see.

This argument also suggests the need for a shift in focus from some accounts that refer more directly to organizations and individuals, to ones that acknowledge – as privacy advocates and others have argued for some time – that online subjects are also difficult to define, are not really amenable to the kinds of individualist characterizations common in some "privacy" discourses and are hard to connect with the kinds of actors that might be called upon to raise questions about Big Data surveillance in the political realm. There exists, of course, a recognition of privacy in human rights codes and constitutional documents but keeping these in view is a constant challenge.

The Deleuzian approach appears to fit the surveillance-analysis bill in some ways, because it acknowledges the shift from just the "state" to other surveillance agencies, from "individuals" to "dividuals" and from discipline to control. As Mattelart and Vitalis observe, a Deleuzian approach also highlights the mobile and invisible nature of much surveillance and the ways that it depends on the involuntary participation of individuals – metadata again – and the purpose of anticipating behavior (Mattelart and Vitalis, 2014) that is overwhelmingly evident in the Snowden disclosures. At the same time, a Deleuzian approach is misleading if one imagines that the world of top-down government-based surveillance is a thing of the past. Such practices now appropriate data from the "rhizomic" forms of surveillance described by Deleuze.

Surveillance in the era of Big Data, then, does not focus only on the body or on a population but on definitions to which we may contribute as part of our daily online interactions. It "makes up" the data double, Deleuze's "dividual" and that entity then acts back on those with whom the data are associated, informing us who we are, what we should desire or hope for, including who we should become. The algorithms grip us even as they follow us, producing ever more information to try to make the user data more effective. Users discover, one might say, that the price of our freedom in both political and consumer contexts is our shaping or conditioning by algorithms.

Anticipation

The political-economic and socio-technical responses to 9/11 helped to change the "tense" of surveillance in some significant ways (Genosko and Thompson, 2006). Since at least the 1990s, risk-management techniques have increasingly turned towards attempts to predict and preempt future developments but the anticipatory approach was racheted up some further notches as early forms of data analytics were brought into play. The frequently advertised notion of "connecting the dots" was predicated exactly on what might be called anticipatory analytics, where the aim of amassing and mining data was "knowledge discovery," of finding patterns in data that would point a suspicious finger towards persons and groups whose associations or communications added up to a "person of interest" profile. In other words, not merely what they might be but what they might be*come*, was a significant factor in assigning riskiness from which it was a short step to suspicion (Kerr and Earle, 2013).

Big Data builds on these already existing modes of anticipatory surveillance in an attempt to create new knowledge using the statistical power of large numbers to help grasp the fragmented details of individual lives. The anticipatory approach is common across the range of Big Data applications. Google Now, for example, uses just this method to draw on a vast concatenation of relatable data in order to alert specific users to things that may have great import for them, from warning them about delayed flights to offering early diagnoses of flu (Regalado, 2013). Everyone collects and transmits much data, especially using smart phones, but also through using any digital device. However, what Lazar calls "Big Data hubris" appears when it is assumed that Big Data – in this case, based on user searches for information about flu – can substitute for rather than supplement conventional modes of analysis. As it happens, the conventional forms of analysis still seem to have a high degree of validity compared with crowd-sourced methods (Butler, 2013). If this is true of epidemiology how much more care should be taken with risk analysis relating to (another rather elastic concept) "terrorism." In this case, as opposed to that of flu, there is no regular presentation of accurate, identifiable, and actionable intelligence. The term itself is politicized, it is well-nigh impossible to distinguish between a violent and non-violent activist, and with so few facts, correcting for false positives and negatives is both rickety and risky.

The situation is exacerbated by the fact that anticipatory approaches are less concerned with the overall picture of a given individual as with "premediating and pinpointing potential dangers" (de Goede, 2014). The problem is that profiles may be built and inferences

made about individuals with privacy regulations and data protection in place. The conventional links between data and the individual have become tenuous and torn. Amoore (2014: 110) asks if privacy rights can be associated with Deleuze's "dividual"? Much filtering and analysis is done, as noted above, *before* identifiable individuals come in sight. She further suggests that the harms are therefore to the " . . . associational life, to the potentiality of futures that are as yet unknowable" and to the very possibility of making a political claim (111).

Science-Technology-and-Society approaches are important for indicating the importance of an ontological approach to the making-up of data subjects but this by no means should condone complacency about the ways that subject positions are still imbricated with neo-liberal notions of, for example, "deserving poor" that still characterize some welfare system use of Big Data practices (Maki, 2011) or the profiling of "bad guys" (why do intelligence and policing agencies persist in using such terms?) in anti-terrorism units. As other studies have indicated, such invidious categories not only persist but are also amplified as greater reliance is placed on automated (see e.g. Gandy, 2012) and actuarial (see e.g. Harcourt, 2007) methods. One has to ask, what do the increasing flows of those data between different kinds of organizations mean for the reproduction of social distinctions – class, gender, ethnicity – and for public accountability of data processing bodies?

Big Data practices also encourage the use of automated decision-making and thus downplay the role of discretion (see also Ruppert). Seen in classic liberal-legal terms, automated decisions can easily deprive individuals of their liberty and property, that trigger in the US the safeguards of the Due Process Clauses of the Fifth and Fourteenth Amendments. For example, computers can terminate individuals' Medicaid benefits, impairing a statutorily-granted property interest (see Citron, 2008). Innocent individuals may be designated as dead-beat parents, resulting in lost property, revoked driver's and professional licenses, and injury to their reputations. The US federal government's "No Fly" data matching program labels some individuals as potential terrorists, resulting in the postponement or denial of air travel, both significant impairments of liberty rights. Automation, suggests Danielle Citron, will be a driving force in the retreat from the discretionary model of administrative law. Nonetheless, due process does mean that citizens or consumers can push back against such automation when it precludes or limits understanding or responding to suspicions, charges, or cut benefits. However, such an assumption depends on those citizens and consumers knowing what is happening, which Big Data approaches make very difficult if not impossible.

Adaptation

We noted earlier that many Big Data practices are common across different platforms. In this section, however, we indicate that what might under some circumstances be acceptable for Google may be highly unacceptable where the NSA is concerned. Google, after all, holds the contents of much of the visible internet in its data centers and this includes satellite images, ground level photos of the built environment in a geospatial database indexed to individuals and organizations. The electronic activities of hundreds of millions of people, including emails and search requests are also known to Google.

As Sean Gallagher (2013) observes, what the NSA does is essentially similar, capturing call metadata and gaining access to information like that of Google through systems like Tempora and possibly PRISM. But the additional factor is that the NSA, using the Foreign Intelligence Surveillance Act, can follow up "exceptions" with warrants to check on persons of interest. This has been possible for some time, a fact first exposed by former AT&T employee Mark Klein in 2005, when he showed how AT&T helped the NSA to gain access to its own systems through a splitter that fed into the Intelligence Traffic Analyzer. It was also shown in 2006 that the NSA used its phone call database for social network analysis and, according to information from Snowden in 2013, call data collection of US to foreign numbers is still occurring.

Curiously, solving just such problems of data storage and analysis have been key to the operations of Google and Yahoo!, which prompted the NSA to improve on Google's BigTable systems with a program called Accumulo, that has multiple levels of security access. It can also generate near real-time reports from data patterns, such as words or addresses from a range of IP addresses, right across the internet. Through what are called "iterators," emergent patterns are constantly reported back to the NSA so that it can "visualize" links between entities based on relationships and attributes. In this way it resembles Facebook's social graph, which is a global mapping system of users and how they are related to each other; it is the largest social network dataset in the world. PRISM offers online NSA access to cloud providers, primarily seeking metadata, which completes the circle. The NSA's new data center in Utah with its huge data-storage capabilities will enable the expansion of PRISM-type real-time internet surveillance (although being classified, the precise purposes are unpublished).

One question for privacy advocates and others is whether or not these surveillance operations are legal: they contend that such programs violate laws designed to protect the liberty and privacy of citizens.

The assurances given, in the US and other countries, that citizens are not targeted by these systems, have failed to reassure citizens and privacy advocates. It is also clear that some data are "incidentally" collected on the whereabouts of domestic cellphones. Such data may be used to map users' relationships, as noted earlier in relation to CoTraveler (see Gellman and Soltani, 2013).

It is crucial to distinguish between different kinds of consequences. As noted above, marketing uses of Big Data analytics cannot simply be extended to anti-terrorist pre-emption. Marketers will be satisfied with results that are accurate only in a relatively small proportion of cases, just because the cluster around that group will also be profitable, albeit to a lesser extent. The economic harms to individuals from such inaccuracy, though potentially serious, are seldom considered by marketers (see e.g. Gandy, 2013; Turow, 2012) and are in some contexts fairly inconsequential (Amazon suggesting some books in which readers have no interest, for example). But the attempt to find terrorist "needles" in Big Data "haystacks" is fraught with palpable problems. Such "needles" are, generally speaking, clever, determined, and imaginative in their attempts to evade detection. The needle-and-haystack argument carries with it a high probability of false positives, which do matter immediately and intensely because the likelihood is high of harm to specific persons.

Critique

The question of Big Data, understood in relation to the Snowden disclosures, has generated unprecedented public interest in surveillance in many countries around the world. While technical and legal responses have been made and while at the level of civil society much activity is evident, particularly demanding accountability – and, where appropriate, abolition of some programs – from the NSA and its cognate agencies, less progress has been made on what might be called a broad ethical front. Yet the questions raised are profound ones for which there are no ready answers and thus, I suggest that an ethical turn becomes more urgent as a mode of *critique*. This is so at several levels, but particularly in the kinds of ways that Snowden himself indicates through his repeated questions about "what kind of society do we want?"

We began with the question of capacity which is reflected in the popular metaphors used about Big Data, notably the handily alliterating "data deluge." The metaphors associated with Big Data are revealing for the hopes and fears associated with Big Data. As Deborah Lupton (2013) has observed, many are associated with liquidity. Unlike the metaphors first adopted for computer technologies, that invoke a "natural" world of the web, cloud, bug, virus, mouse, and spider, Big Data tropes "relate to streams, flows, leaks, rivers, oceans, waves" but also to floods or tsunamis that may seem to threaten to swamp or drown us. They are potentially uncontained, out-of-control. But there is more to the liquidity issue than metaphors such as the data deluge.

Under the heading "liquid surveillance" I discussed with Zygmunt Bauman the ways in which data flow increasingly freely within and between containers and in particular the ways that digital surveillance has a seemingly symbiotic relationship with the kind of liquidity visible in contemporary social, political, and economic arrangements, that are often short-term, fissiparous (Bauman and Lyon, 2013). They also query the kinds of "blockages and resistances, the solidities that may impede the fluid circulation of data" (Lupton, 2013) that tend to be omitted from the free-flow-of-data accounts. The liquidity of surveillance is as significant in the social and political realm as at the level of data-flows.

One theme of *Liquid Surveillance* is the need for properly ethical practices. Big Data is currently dominated by commercial and governmental criteria and these are often met with technical demands (for better encryption for example) or legal demands (for legislation relevant to today's technologies). Privacy advocates and internet activists also try to promote new political approaches to emergent tendencies such as Big Data. But a key reason why those commercial and governmental criteria are so imbricated with Big Data is the strong affinity between the two, particularly in relation to surveillance. Big Data represents a confluence of commercial and governmental interests; its political economy resonates with neo-liberalism. National security is a business goal as much as a political one and there is a revolving door between the two in the world of surveillance practices (Ball and Snider, 2013).

Properly ethical practices are at a relative disadvantage for several other reasons as well. Not many ethicists spend time thinking about the complexities of the internet, social media, or Big Data and many of those at the forefront of the Big Data field seem to have little time for ethics except as a minor, residual concern (see Narayanan and Vallor, 2014). The imperatives for Big Data approaches come from a belief in the immense power of technology – can Google really track and predict the spread of flu faster than centers for disease control? (Ginsberg et al., 2009; Lazer et al., 2014) – along with the capacity to analyze vast quantities of data at steadily shrinking unit costs. But just as in the Google flu example, questions must be asked about how good are the surveillance data and the modes of analysis?

How data are generated and framed always has decisive effects on the final outcomes of analysis. As Lisa Gitelman reminds us, ''raw data is an oxymoron'' (2013); data have always been ''cooked'' as Geoff Bowker says in the conclusion of Gitelman's book. Terms such as metadata, so crucial to Big Data surveillance, lack clear definition, even though it can generally be distinguished from data such as the content of phone calls or emails. Yet those ill-defined metadata are used, constantly, by security and intelligence agencies, and the patterns revealed by the algorithms used to filter them relate back to the purposes that shape the data in the first place and forward to those affected by the designation of groups that may contain persons of interest.

The range of ethical issues relating to Big Data surveillance is considerable, but from what has been discussed in the foregoing, may be clustered as *privacy*, *social sorting*, and *preemption*.

Given the reliance on western liberal legal traditions it is hardly surprising that public debate generally commences around the question of *privacy*. Understood as a human right, it underlies aspects of democratic polity, such as freedom of expression. Often understood in the post-Snowden era as relating to control of communications about oneself, it is clearly a threatened value if not – according to some – a forlorn hope. Following the above argument, though, it is vital that an ethics of Big Data practices be found that deals with the problem of the increasing gap between data and individuals (Amoore, 2014; Stoddart, 2014). But as privacy is still the preeminent mobilizing concept for opposition to inappropriate, disproportionate or illegal surveillance, the efforts of those who propose technical limits such as encryption or de-identification or who would re-infuse the concept with content appropriate to a Big Data world are certainly welcome.

As far as *social sorting* is concerned, this is a concept that alerts us to several related practices that produce uneven and unequal outcomes when the supposedly neutral and illuminating techniques of Big Data – especially predictive profiling – are applied to perceived social and political problems. This connects surveillance both with modern bureaucratic practices and also, under the sign of security, with insurance logics that see security as procurable through intelligence gathering, identification, and tracking (Lyon, 2007; Zedner, 2009). Its outcomes – amplified in Big Data contexts – are above all the growth of categorical suspicion (the parallel in consumer surveillance, is what I term ''categorical seduction'' Lyon, 2007). This in turn encourages a consequentialism that departs from earlier notions of proportionate punishment to deterrence and incapacitation. Together with a ''penal populism'' that calls for public protection, reinforced by media-enhanced perceptions of risk, time-honored commitments to the presumption of innocence, or proof beyond reasonable doubt are eroded (Zedner, 2009: 80).

Thirdly, an emphasis on *preemption* takes the actuarial logic one stage further, connecting with what was said above about how Big Data fosters an anticipatory, future tense approach to surveillance. Again this is not a new development in surveillance. Risk-management in particular has encouraged such anticipatory governance for several decades. But the availability of Big Data techniques encourages an intensified future-orientation in practice. So the possibility that, because of certain data fragments, the data-body may be thought to have a propensity to certain behaviors that are not yet evident, leads to some action. The data have effects; they are, as Rita Raley says, ''performative.'' Following Haggerty and Ericson's (2000) Deleuzian discussion of the surveillant assemblage, Raley points out that,

> the composition of flecks and bits of data into a profile of a terror suspect, the re-grounding of abstract data in the targeting of an actual life, will have the effect of producing that life, that body, as a terror suspect. (Raley, 2013: 128)

Conclusion

The main question addressed in this article is in two parts: One, in what ways and to what extent do the Snowden disclosures indicate that Big Data practices are becoming increasingly important to surveillance? The answer, clearly, is yes, they are. Many of the major Snowden revelations, especially those in which metadata feature prominently, indicate a reliance upon Big Data practices. The second question, following on from the first, is how far does this indicate changes in the politics and practices of surveillance? Are new trends, or the augmentation of older ones, visible here? Again, the evidence discussed here suggests strongly that Big Data practices are skewing surveillance even more towards a reliance on technological ''solutions,'' and that this both privileges organizations, large and small, whether public or private, reinforces the shift in emphasis towards control rather than discipline and relies increasingly on predictive analytics to anticipate and preempt.

These questions were explored in respect to the *capacities* of Big Data, their social-political *consequences* and the kinds of *critique* that may be appropriate for assessing and responding to these developments. For the first, I argue that ''size'' is not directly the issue but rather that, taken together, the loose cluster of attributes of ''Big Data'' make a difference in ways that are hard to generalize. Big Data practices echo several key surveillance trends but in several respects

they point to realities that have perhaps been underestimated. One is that, within surveillance studies there has been a general tendency to analyze multiple forms of surveillance that are not directly linked with state-based, top-down surveillance of the kind epitomized in George Orwell's *Nineteen-Eighty-Four*. If this was understood by some to mean that more generalized – or, following Gilles Deleuze, "rhizomic" – surveillance spells less state surveillance activity, the Snowden revelations are rapidly dispelling that illusion.

However, those revelations, which as I show above, indicate an increasing dependence on Big Data practices, also lay bare in ways that were known only hazily before just how far security and intelligence agencies depend on data obtained from the commercial realm. These are *consequences* that cry out for careful consideration. In a sense, this means that Orwell's bleak vision of what tendencies in post-war liberal democratic polities could lead to authoritarian surveillance regimes were not mistaken so much as standing in need of complementary analyses, such as that of his contemporary, Aldous Huxley, in *Brave New World*. Big Data practices in consumer surveillance are (now literally!) co-travelers with those of state surveillance and together produce the kinds of outcomes around which ethical debates should now revolve. Indeed, not only are they "co-travelers," they also cooperate extensively, the one taking methods from the other, with, as discussed above, potentially pernicious results as the "successful" methods in one area are applied in ways deleterious of human rights in another. Sadly, little time seems to be spent on such matters in typical computing studies departments in today's universities, where all too often notions like privacy and civil liberties are regarded as a nuisance that slows research development (Narayanan and Vallor, 2014).

It is these matters in particular that attract *critique*, especially in relation to anticipatory and preemptive approaches common to Big Data mindsets and activities and amplifying what is a long-term surveillance trend. These fit neatly, of course, with currently intensifying political styles of neo-liberalism that, with regard to "national security," are seen in a list towards actuarialism and a consequentialist concern with managing disorder and crime rather than seeking its causes and attempting to eradicate them (Agamben, 2013). Let me give two examples. Critically, certain time-honored legal protections such as a presumption of innocence or proof beyond reasonable doubt are being eroded within a number of western societies precisely due to the developing reliance on big-data-led beliefs that suspects can be isolated by category and algorithm. Even if one-time "suspects" have their names cleared by judicial process, the fact that Big Data practices exemplified in the collect-it-all slogan include retaining data indefinitely, it can be hard for persons with a "record" ever to make a fresh start. Data in the Canadian Police Information Centre, for example, remain there permanently. And when police include mental health problems in their records these can lead to denial of entry to Canadians trying to cross the border into the US. Attempted suicide calls, for example, have been uploaded to international databases with just this outcome (CBC, 2014).

Snowden's revelations have done good service in showing how far state-based surveillance extends but also how much it depends on Big Data practices that implicate corporate bodies and connect directly with everyday practices of ordinary internet and cellphone users. Ethically, he frequently, and wisely, asks what kind of society we want to live in. Is it one marked by fear and mutual suspicion, where data are collected promiscuously and kept forever, in systems that never forget, making forgiveness obsolete and creating much to fear even though you have nothing to hide? Is it one where vulnerability is amplified, democracy diminished and where ordinary people are more exposed to organizations that are themselves more opaque? These are questions that Big Data surveillance obliges us to confront.

Acknowledgement

The author wishes to thank Chris Parsons and Ian Kerr and also two anonymous reviewers for their constructive critique.

Declaration of conflicting interest

The author declares that there is no conflict of interest.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Notes

1. https://www.schneier.com/blog/archives/2013/09/metadata_equals.html/

References

Agamben G (2013) For a theory of destituent power. *Critical Legal Thinking*. Available at: http://criticallegalthinking.com/2014/02/05/theory-destituent-power/ (accessed 19 June 2014).

Amoore L (2011) Data derivatives: On the emergence of a security-risk calculus for our times. *Theory, Culture and Society* 28: 24–43.

Amoore L (2014) Security and the claim to privacy. *International Political Sociology* 8(1): 108–112.

Andrejevic M and Gates K (2014) Big Data surveillance: Introduction. *Surveillance & Society* 12(2): 185–196.

Ball KS and Snider L (eds) (2013) *The Surveillance-Industrial Complex: A Political Economy of Surveillance*. London: Routledge.

Bankston KS and Soltani A (2014) Tiny constables and the cost of surveillance: Making cents out of the United States vs Jones. *Yale Law Journal Online*. January 09. Available at: www.yalelawjournal.org/the-yale-law-journal-pocket-part/constitutional-law/tiny-constables-and-the-cost-of-surveillance:-making-cents-out-of-united-states-v.-jones/ (accessed 19 June 2014).

Bauman Z and Lyon D (2013) *Liquid Surveillance: A Conversation*. Cambridge: Polity Press.

Bennett CJ, Haggerty K, Lyon D, et al. (eds) (2014) *Transparent Lives: Surveillance in Canada*. Edmonton: Athabasca University Press.

Bertolucci J (2013) Big Data's new buzzword: Datafication. *Information Week*. February 25. Available at: www.informationweek.com/big-data/big-data-analytics/big-datas-new-buzzword-datafication/d/d-id/1108797?/ (accessed 19 June 2014).

Boyd D and Crawford K (2012) Critical questions for Big Data: Provocations for a cultural, technological and scholarly phenomenon. *Information, Communication and Society* 15(5): 662–679. Available at: www.tandfonline.com/doi/abs/10.1080/1369118X.2012.678878#.UthCMvZA_EV/ (accessed 19 June 2014).

Butler D (2013) When Google got flu wrong. *Nature*. 494(7936). Available at: www.nature.com/news/when-google-got-flu-wrong-1.12413/ (accessed 19 June 2014).

CBC (2014) Canadians' mental health info routinely shared with FBI, U.S. Customs. Available at: www.cbc.ca/news/canada/windsor/canadians-mental-health-info-routinely-shared-with-fbi-u-s-customs-1.2609159/ (accessed 19 June 2014).

Cheney-Lippold J (2011) New algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture and Society* 28(6): 164–181.

Citron D (2008) Technological due process. *Washington University Law Review* 85(6): 1249–1313.

de Goede M (2014) The politics of privacy in the age of preemptive security. *International Political Sociology* 8(1): 100–104.

Deleuze G (1992) Postscript on the societies of control. *October* 59: 3–7.

FAS (Federation of American Scientists) (2014) Available at: www.fas.org/irp/budget/index.html?PHPSESSID=70809e6b347db7b2122df1ef24d743e0/ (accessed 19 June 2014).

Freeze C (2014) Canada's metadata collection worries critics. *The Globe and Mail*. March 27. Available at: http://m.theglobeandmail.com/news/world/canadas-metadata-collection-worries-critics/article17714407/?service=mobile/ (accessed 19 June 2014).

Gallagher S (2013) What the NSA can do with "Big Data." *Ars Electronica*. June 11. Available at: http://arstechnica.com/information-technology/2013/06/what-the-nsa-can-do-with-big-data/ (accessed 19 June 2014).

Gandy O (1993) *The Panoptic Sort: A Political Economy of Personal Information*. Boulder, CO: Westview Press.

Gandy O (2012) Statistical surveillance: Remote sensing in the digital age. In: Ball KS, Haggerty K and Lyon D (eds) *Routledge Handbook of Surveillance Studies*. London and New York: Routledge, pp. 125–132.

Gandy O (2013) *Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage*. London: Ashgate.

Gellman B and Poitras L (2013) US, British intelligence mining data from nine US internet companies in broad secret program. *The Washington Post*. June 6. Available at: www.washingtonpost.com/investigations/us-intelligence-mining-data-from-nine-us-internet-companies-in-broad-secret-program/2013/06/06/3a0c0da8-cebf-11e2-8845-d970ccb04497_story.html (accessed 19 June 2014).

Gellman B and Soltani A (2013) NSA tracking cellphone locations worldwide, Snowden documents show. *The Washington Post*. December 4. Available at: www.washingtonpost.com/world/national-security/nsa-tracking-cellphone-locations-worldwide-snowden-documents-show/2013/12/04/5492873a-5cf2-11e3-bc56-c6ca94801fac_story.html?hpid=z1/ (accessed 19 June 2014).

Genosko G and Thompson S (2006) Tense theory: The temporalities of surveillance. In: Lyon D (ed.) *Theorizing Surveillance: The Panopticon and Beyond*. Cullompton: Willan, pp. 123–138.

Ginsberg J, et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457: 1012–1014. Available at: www.nature.com/nature/journal/v457/n7232/full/nature07634.html/ (accessed 19 June 2014).

Gitelman L (ed.) (2013) *Raw Data is an Oxymoron*. Cambridge, MA: MIT Press.

Glennon NJ (2014) National security and double government. *Harvard National Security Journal* 5(1). Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376272 (accessed 19 June 2014).

Greenwald G (2013) NSA collecting phone records of millions of Verizon customers daily. *The Guardian*. June 6. Available at: www.theguardian.com/world/2013/jun/06/nsa-phone-records-verizon-court-order/ (accessed 19 June 2014).

Greenwald G, MacAskill E and Poitras L (2013) Edward Snowden: The whistleblower behind the NSA surveillance revelations. *The Guardian*. June 10. Available at: www.theguardian.com/world/2013/jun/09/edward-snowden-nsa-whistleblower-surveillance/ (accessed 19 June 2014).

Haggerty K and Ericson R (2000) The surveillant assemblage. *The British Journal of Sociology* 51(4): 605–622.

Harcourt B (2007) *Against Prediction: Profiling, Policing and Punishing in an Actuarial Age*. Chicago: University of Chicago Press.

Kerr I and Earle J (2013) Prediction, preemption, presumption: How Big Data threatens big picture privacy. *Stanford Law Review* 66(65). Available at: www.stanfordlawreview.org/online/privacy-and-big-data/prediction-preemption-presumption/ (accessed 19 June 2014).

Kitchin R (2014) Big Data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267.

Kitchin R (forthcoming) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.

Lanchester J (2013) The Snowden files: Why the British public should be worried about GCHQ. *The Guardian* October 3. Available at: www.theguardian.com/world/2013/oct/03/edward-snowden-files-john-lanchester/ (accessed 19 June 2014).

Lazer D, Kennedy R, King G, et al. (2014) The parable of Google flu: Traps in Big Data analysis. *Science* 343: 1203–1205. Available at: www.sciencemag.org/content/343/6176/1203.full/ (accessed 19 June 2014).

Lupton D (2013) Swimming or drowning in the data ocean? Thoughts on the metaphors of Big Data. Available at: http://simplysociology.wordpress.com/2013/10/29/swimming-or-drowning-in-the-data-ocean-thoughts-on-the-metaphors-of-big-data/ (accessed 19 June 2014).

Lyon D (2001) *Surveillance Society: Monitoring Everyday Life*. Berkshire: Open University Press.

Lyon D (2007) *Surveillance Studies: An Overview*. Cambridge: Polity.

Maki K (2011) Neoliberal deviants and surveillance: Welfare Recipients under the watchful eye of Ontario Works. *Surveillance & Society* 9(1/2): 47–63. Available at: http://library.queensu.ca/ojs/index.php/surveillance-and-society/article/view/deviants (accessed 19 June 2014).

Mattelart A and Vitalis A (2014) *Le Profilage des Populations: Du Livret Ouvrier au Cybercontrôle*. Paris: La Découverte.

Meyer-Schönberger V and Cukier K (2012) *Big Data: A Revolution that will Transform How we Work, Think and Live*. New York: Mariner.

Mosco V (2014) *To the Cloud: Big Data in a Turbulent World*. New York: Paradigm.

Narayanan A and Vallor S (2014) Why software engineering courses should include ethics coverage. *Communications of the ACM* 57(3): 23–25.

Newell BC (forthcoming) The massive metadata machine: Liberty, power and mass surveillance in the U.S. and Europe. *I/S: A Journal of Law and Policy for the Information Society*.

Raley R (2013) Dataveillance and countervailance. In: Gitelman L (ed.) *Raw Data Is an Oxymoron*. Cambridge, MA: MIT Press, pp. 121–145.

Regalado A (2013) The data made me do it. *Technology Review* May 03. Available at: www.technologyreview.com/news/514346/the-data-made-me-do-it/ (accessed 19 June 2014).

Richards NM and King J (2013) Three paradoxes of Big Data. *Stanford Law Review Online* 66(41): 41–46.

Ruppert E (2012) The governmental topologies of database devices. *Theory, Culture and Society* 29(4–5): 116–136.

Savage M (2013) Digital fields, networks and capital. In: Orton-Johnson K and Prior N (eds) *Digital Sociology: Critical Perspectives*. Houndmills, UK: Palgrave Macmillan.

Savage M and Burrows R (2007) The coming crisis of empirical sociology. *Sociology* 44(5): 885–899.

Schneier B (2012) *Liars and Outliers: Enabling the Trust that Society needs to Thrive*. New York: Wiley.

Schneier B (2014) CSEC analysis of IP and user data. Available at: www.schneier.com/blog/archives/2014/02/csec_surveillan.html/ (accessed 19 June 2014).

Stoddart E (2014) (In)visibility before privacy: A theological ethics of surveillance as social sorting. *Studies in Christian Ethics* 27(1): 33–49.

Trottier D (2012) *Social Media as Surveillance*. London: Ashgate.

Turow J (2012) *The Daily You: How the New Advertising Industry Is Defining Your Identity and Your Worth*. New Haven, CN: Yale University Press.

van Dijck J (2014) Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12(2). Available at: http://library.queensu.ca/ojs/index.php/surveillance-and-society/article/view/datafication/ (accessed 19 June 2014).

White House (2014) Big Data and the future of privacy. Available at: www.whitehouse.gov/blog/2014/01/23/big-data-and-future-privacy/ (accessed 19 June 2014).

Zedner (2009) *Security*. London and New York: Routledge.