BEYOND BIGNESS: CAN BIG DATA HAVE AN ETHICAL FUTURE?

By Tyne Daile Sumner

During occasions when new industries and new technologies are developed, the physical and physiological details usually taken as given can become a matter of concern with consequent clarification of the assumptions and conceptions we have of what individuals are.

—Erving Goffman (1969, p. 4)

ABSTRACT

Questions of privacy and security have long been associated with the collection and use of big data. Increasingly, however, critics have come to associate big data with concepts such as fairness, accountability and transparency. As the scale and complexity of data continues to expand, scholars are being called upon to offer up new ways of tackling the complex ethical debates at the centre of big data. Research presented here covers a range of issues that cut across social, political and technological applications to consider the effects of large-scale data in the twenty-first century. By asking 'Who is missing?' in big data, we can, by extension, consider questions such as: 'What are the ethical debates that matter?' and 'What still needs to be considered beyond the basic codes and protocols for the governance of data?.' Moreover, one of the first necessary steps in rethinking big data's ethical future is a reconceptualisation of the very notion of 'bigness.'

INTRODUCTION

The collection and use of data is unquestionably one of the central preoccupations of our time. More recently, big data has come to dominate social and political fields as diverse as medicine, business, advertising, social media and the news. Described by Mayer-Schönberger and Cukier (2013, p. 6) as "things one can do at a large scale that cannot be done at a smaller one," big data has also begun to have profound effects on the ways we see ourselves as citizens of the world, such that it now clearly shapes our subjectivities (Schroeder, 2018, p. 127).[4] Labelled by some scholars as 'datafication,' this worrying trend can be described as the "quantification of social interaction and their transformation into digital data" (Richiterich, 2018, p. 1). Via this formulation, not only are our tastes, preferences and choices affected by data, the way in which we relate to other people and articulate ourselves is intimately connected to big data's influence. As renowned Information Studies scholar Christine Borgman reminds us, however, "big data is not necessarily better data." She notes the way in which, the farther the observer is from the point of origin, "the more difficult it can be to determine what those observations mean - how they were collected; how they were handled, reduced, and transformed; and with what assumptions and what purposes in mind" (2015, p. xvii).

Comments such as this should alert us to the fact that any system of data collection and use is always bound to a framework—either it cultivates inclusion or, failing that, it systematically works to exclude low-income, minority or underserved communities from, for example, access to society's broader benefits. Moreover, as Eubanks (2018, p. 7) points out in the revelatory *Automating Inequality: How high-tech Tools Profile, Police and Punish the Poor:*

[M]arginalized groups face higher levels of data collection when they access public benefits, walk through highly policed neighbourhoods, enter the health-care system, or cross national borders. That data acts to reinforce their marginality when it is used to target them for suspicion and extra scrutiny. (see also Browne, 2015; Lyon, 2003; Mann & Daly, 2018; Noble, 2018; Gangadharan, 2012; and Sandvig, Hamilton, Karahalios & Langbort, 2016).

While some forms of data-driven inequality arise via the removal of certain people or voices from a collective, others employ digital surveillance technologies to over-monitor particular people or groups of people, usually to damaging effect (Ferguson, 2017). This article gathers a range of recent reflections in the field, alongside a short history of the concept of privacy, to suggest new ways of tackling the complex ethical debates at the center of big data. The relevance of concepts such as justice, empathy, agency, ownership, privacy, subjectivity and identification are also of considerable importance to the disscusion that follows.

4. Schroeder (2018, p. 127) offers another dimension to this definition, noting: "'Big data' can be defined as research that represents a step change in the scale and scope of knowledge about a given phenomenon."

BIG DATA: HOW DID WE GET HERE?

Of the three 'v's' that have come to define big data—volume, variety and velocity—it is perhaps volume that has attracted the most attention.[5] Most information that was formerly stored in wallets or filing cabinets is now digital and growing at an accelerating pace. This is a technological development that has afforded unprecedented data access to more people than ever before. Yet while there is an almost global understanding that personal data should be protected, an individual's private information is nevertheless susceptible to the same exploitative systems that have historically infiltrated other social phenomena: trade, economics, politics, education, climate and so on. Of course, all of these things are inextricably connected to big data, yet somehow the narrative has prevailed that the mass collection of an individual's private information sits outside the realm of the day-to-day functioning of a society.

One way out of this paradox is to foreground the fact that big data refers not only to data, as such. Rather, it encompasses the behaviours, practices, networks, infrastructures and politics that influence and are influenced by its manifestations. Understanding these overlaps is one way of understanding big data as a set of "emerging technology" practices since it also encompasses "digitally enabled developments in data collection, analysis, and utilisation" (Richterich, 2018, p. 23). Moreover, one of the reasons why big data has been increasingly tied to broad debates about human rights, autonomy, transparency, privacy, security and self-responsibility is because it fundamentally challenges pre-existing moral and ethical norms. It does this by advancing the cumulative knowledge of data-collecting organisations, thereby also advancing the power gained over individuals and groups. A now widely-understood offshoot of this trend is the "application of big data knowledge in shaping media uses," which thereby has dramatic effect on the social implications of such media usage (Schroeder, 2018, p. 127). Big data poses a challenge also to pre-existing ethical frameworks with regards to consent, insofar as corporate data economies have succeeded in organising big data's alleged lack of bias towards commercial gain. It is for this reason that any intellectual inquiry into big data requires a more robust methodology than simply probing the data itself; it also necessitates an interrogation of the knowledges and power structures that underpin its collection and usage in the first instance. In order to address this issue, many scholars have come to see big data as an overarching framework for understanding the contemporary technological landscape. In *We Are Data: Algorithms and the Making of our Digital Selves,* Cheney-Lippold (2017, p. 4) notes, for example, the ways in which our "daily activites are mediated with software" such that the "resulting aggregation of our lives' data founds the discursive terrain of our digital environments." Similarly, others note the way in which big data functions as a crucial "sense-making" resource in the digital era (Andrejevic, 2014, p. 1675).

While these are useful ways of thinking about big data's ubiquity and dominance, these datasets nevertheless lend themselves to problematic misinterpretation (Harford, 2014). One way in which the pervasiveness of big data is prone to being misunderstood is in the assumption that size and scale somehow equate to lack of bias. The common coinage 'digital positivism' goes some way towards explaining this assumption insofar as it encapsulates a range of theories that assert that data—in ways similar to the physical world—operates according to general or universal laws (Mosco, 2015 & 2016). We need look no further, however, to the now common example of big data policing—euphemistically known as 'predictive reasonable suspicion'—to know with certainty that more data does not necessarily correlate to more ethical data systems. With more specific information, police officers may now be afforded a stronger predictive sense that they are in fact observing a criminal act. As Ferguson (2015, p. 331) points out, however, the "next phase will use existing predictive analytics to target suspects without any firsthand observation of criminal activity, relying instead on the accumulation of various data points." The underlying problem with this formulation is that the very data used for predictive purposes contains built-in sociodemographic bias. Or, to borrow from Ferguson (2015, p. 331) again, "this new reality simultaneously undermines the protection that reasonable suspicion provides against police stops and potentially transforms reasonable suspicion into a means of justifying those same stops."

One way in which big data holders and organisations have attempted to defend against this reality is by asserting the relevance of informed consent. While there exists an abundance of material on 'best practices' for informed consent in relation to data—especially for enabling the reuse of research data beyond the purpose for which it was collected—the fuzzy ethics surrounding informed consent cannot be ignored (Koops, 2014; Gellert & Gutwirth, 2013; and Parsons, 2015). Part of this problem is with regards to transparency, while there is also the complicating factor of limited citizenry knowledge of how big data actually operates. "Often, when confronted with the potential of using personal data," Matzner (2014, p. 96) comments, "people react surprised and affected suggesting that they would not have consented to this use of their data if they had been informed about the possible consequences." This is further complicated by the fact that even in situations where it is possible to acquire comprehensive information about the nature, collection and use of data, the complexity of the process and effort involved poses ethical problems in itself (van der Ploeg, 2007, p. 49). As Matzner (2014, p. 96) goes on to assert, "it is questionable if such an effort can reasonably be required by everybody or whether this establishes new inequalities in terms of knowledge and skills necessary to use a service or technology." Thus, the bigness of big data in turn creates a situation whereby processes of comprehension or discernment itself can generate new and potentially unequal power structures.

5. Note that some scholars have sought to add a fourth 'v'—veracity—to this paradigm, in order to draw attention to questions of reliability around certain data usage.

Recent personal data breaches by Facebook, for example, have drawn worldwide attention to the problematic and slippery frameworks that underpin the data collection processes for many large corporations. Following widespread uproar over the Cambridge Analytica data leak, Facebook now restricts developer access to user data. It is a move that Schroepfer (2018, para. 1), Facebook's chief technology officer, described as a change that will "better protect people's information while still enabling developers to create useful experiences." Yet while protecting user data from potentially exploitative developers and data-hungry apps might seem like an ethical response to Facebook's privacy problems, the reverse is in fact the case. By restricting access to various Application Programming Interfaces (APIs) and thereby reducing data transparency, Facebook has instead successfully leveraged a massive security data breach to make it harder for outside groups (reseachers, for example) to gain insight into its algorithmic objectives. While this recent protective measure can work on the one hand to block access by developers to users' religious preferences, it also operates on the other to prevent research into Facebook's targeted advertising processes.

As recent events have revealed, these data organising principles are often closely connected to the overarching ideology behind large businesses or corporations. The collection of big data and its associated algorithms reflect broad conceptions of power that are akin to Foucault's (1975) influential model in which power is less a force exerted on individuals and rather a dynamic deeply embedded within societies at large. Facebook's data breach is just one example that exposes the complex entanglements between consent, ethics and corporate big data practices; as we are becoming increasingly aware, there will be more of the same to come (Kramer, Guillory, & Hancock, 2014).

Moroever, the well-known Facebook case highlights a common grey area when it comes to informed consent in the context of big data (PrivazyPlan, 2018).[6] Ultimately, the approval by users of social media privacy policies does not seem sufficient enough grounds for corporations to justify the use of personal data for any variety of commercial or research purposes; rather there should be some clear limitations put in place to further protect these practices (Rothstein & Shoben, 2013; Ioannidis, 2013). In unpacking these views and others, there is also need to consider the role of algorithmic bias; the way it reflects not just the techno-corporate contexts in which the majority of big data is created, but also the political and educational frameworks and organisations through which this data moves and is governed. While the open data movement promotes the accessibility of data as a public good, not all data is created equal nor do all citizens have equal access to it. As Richterich (2018, p. 40) has usefully written in relation to this point, "the 'big data divide' implies power/knowledge conditions that systematically exclude individuals from access to data which would allow them to assess the data generated by corporations, the conditions under which this is done, and how this information is used." This is indeed a key barrier to effectively tackling algorithmic bias at a deeper level regarding unequal access to data in the first place (Powles & Nissenbaum, 2018).[7]

PRIVACY: IS THERE ANY LEFT?

Placing some of the more pressing concerns associated with big data in the context of privacy's long-term erosion might shed some light on whether citizens' concerns have historically had any effect on the organising principles of those who collect and use individuals' data. Afterall, widespread concern over privacy is in no way unique to the current moment. In their seminal essay *The Right to Privacy*, Warren and Brandeis (1890, p. 205) pronounced that the right to privacy was based on a principle of "inviolate personality," thus laying the foundation for the modern understanding of privacy as control over one's personal information. Later, Westin (1967, p. 7) defined privacy as the "claim of individuals, groups or institutions to determine for themselves when, how and to what extent information about them is communicated to others." By the second half of the twentieth century, anxiety around loss of privacy was pervasive. This anxiety was generated in part by new visual and audio technologies, as well as changes in constitutional privacy laws; it was largely due, however, to the intensification of surveillance activity in the early years of the postwar period. A surge of writing emerged in response to such trends, which aimed to not only highlight the large-scale collection of individuals' information by government and corporate bodies, but also the need for a collective ideological resistance to such trends. Dash's 1959 publication The Eavesdroppers opened the gate for an outpouring of texts that examined privacy through an unprecedented sociological lens, as a topic requiring urgent critical attention. Subsequent texts such as Ernst's *Privacy: The Right to Be Let Alone* (1962), Brenton's *The Privacy Invaders* (1964), Westin's *Privacy and Freedom* (1967), and Smith's *Privacy: How to Protect What's Left of It* (1979) represent a snapshot of the period's intense focus on the problems associated with the rapid erosion of personal privacy. Collectively, these texts signal that by the end of the twentieth century, citizens the world over were beginning to accept that the boundaries between their private and public selves were no longer secure. An argument common to many of these publications is that privacy as a concept is, by its very nature, linked to notions of personhood and self-identity (in Kulhari, 2018). The right to privacy—also known as informational self determination—is an "important facet of the right of personality, which guarantees every individual the possibility to develop her own personality" (Kulhari, 2018, p. 28). While current practices of handing over personal data are frequently indirect and ancillary—one

6. While I am using the concept of 'informed consent' in the context of this discussion, it is worth noting that consent is not the only legal basis for processing information (although it depends on jurisdiction). For example, the EU General Data Protection Regulation (GDPR) sets forth six conditions for the lawfulness of processing data: "consent; for the performance of a contract; for compliance with a legal obligation; to protect the vital interests of the data subject or of another natural person; for the performance of a task carried out in the public interest; and for the purposes of the legitimate interests pursued by the controller or by a third party" (PrivazyPlan, 2018).
7. Powles and Nissenbaum (2018, para. 7) consider the extent to which trying to 'fix' AI and algorithmic bias actually distracts from the more urgent questions about the underlying technology used in these systems, as well as the unequal power structures that underpin the data that comprises them in the first place.

example being the divulging of specific tastes and preferences via Social Media platforms—earlier narratives were ones in which a person's subjectivity was not yet modulated or externalised via big data. Despite these changes, the protection of personal data within democratic societies today still tends to be considered an extension of the right to privacy, despite arguments that they are distinct. Scholars who justify the inextricable connection between privacy and data protection frequently foreground claims for the right to data protection being characterised by strong links to the right to privacy (Gonzalez Fuster, 2014). Those who argue that privacy and data protection rights are substantially distinct, frequently invoke the scope and size of each respective category, arguing that although the two often overlap, there are instances of data processing that have nothing to do with personal privacy (Gellert et al., 2013, p. 525).

Increasingly, privacy has less to do with the ways in which individuals choose to disclose personal information, and more to do with the ways in which they interact either directly or indirectly with a wide array of social, political and cultural phenomena. Because big data operates most successfully within overlapping realms of public and private, these practices seem to somehow elude many prewar classifications of subjectivity. Prior to the rise of mass electronic surveillance during the pre-Internet era of the mid-to-late twentieth century, a person's sense of who they were, together with what status their personal data occupied, was relatively unimpeded by technological frameworks. To put this another way, an individual's personal data previously existed in a comprehensible form. As this scenario changed, so too did our collective understandings and expectations of how much autonomy over personal data an individual could and should have. Today, the rise of big data has put the very idea of an individual's self-hood and autonomy under direct threat. Big data critics have reflected this by emphasising the lack of control, knowledge and agency that individuals have over the ways in which their personal information is being collected in relationship to the use of online services (Tene & Polonetsky, 2012). There is also concern related to the insistence of service providers that a user's personal data remains anonymous; an assurance which critics have come to see as almost impossible. Richterich (2018, p. 38) has summarised this concern, where she states that:

Big data enforce an increased, though neither necessarily deliberate nor conscious transparency of online users/consumers. The full extent of this transparency is only visible to those actors controlling the main data collecting platforms or gaining external access to these. What is ultimately collected here, are vast amounts of personal information concerning individuals' preferences, attitudes, moods, physical features and […] health status and health-relevant behaviour.

Questions surrounding data collecting platforms often highlight related privacy violations when it comes to the reidentification of an individual's data, in particular, data stored in either a private or public health record. Public health data exists at the crossroads of several big data tensions, offering useful examples for thinking through the interdependencies of big data practices with forms of health surveillance, scientific research and ethics. Indeed, when we begin to think about the relationship between privacy and processes of data reidentification, several questions come to mind. In the first instance, what does it *mean* for a platform to have 'stored' personal data? Second, what are the implications of frameworks that seek to re-identify data that has previously been de-identified? And finally, do data collecting organisations have an ethical and/or moral responsibility to notify individuals whose data has been intentionally or accidentally leaked? An ostensible 'quick fix' to some of these questions has been the suggestion that banning deidentification practices at the outset of a data collection process would prevent the potential for any subsequent breach. Such drastic measures undoubtedly generate further problems by preventing people who are trying to gain transparency around particular data systems from in-turn interrogating them. Thus, criminalising the reidentification of an individual's data, in any context, might provide enhanced citizen confidence and surety in the short term, but does not actively make the system more ethical in the long term.

ETHICS: WHO IS MISSING AND HOW?

By virtue of its 'bigness,' big data ultimately fosters a culture of what might be called data noise or data saturation. Indeed, critics have recognised this trend as early as the mid-twentieth century, albeit as a product of technologies such as television—now viewed as almost benign in comparison to something like Facebook's election-fixing algorithmic capabilities. As early as the 1950s in America, for example, new forms of media, combined with the corporatisation of modes of communication, were creating a culture in which "public discourse belong[ed] entirely to the mass media, particularly electronic media, [to] include only the voices of those who could penetrate or manipulate a genre of discourse that thrive[d] on overcommunication" (Doreski, 1999, p. 75). Within this arrangement, those who have the power to shape and control the flow of data also have the capacity to prioritise particular narratives. While the machinations and effects of online news and advertising algorithms are the source of much recent critical attention, the social and political implications of big data's tendency to enact processes of exclusion and, conversely, over-inclusion still requires more consideration. The Data Justice Lab at Cardiff Univeristy in the UK represents a key organised approach to articulating these implications, via its development of a research agenda focused entirley on examining the complex relationship between datafication and social justice. The lab maintains, for example, a Data Harm Record, which runs a continual log of problems associated with

automated and algorithmic systems reported from across the globe (Redden, 2018, para. 4). The record divides the broad concept of 'data harms' into eight useful cateogries: commercial uses of data (potentials for exploitation); discrimination; loss of privacy; identity theft, blackmail, reputational damage and/or distress; physical injury; political uses of data, political manipulation and social harm; Government uses of data (data errors); and harms due to algorithm/machine bias (Redden & Brand, 2018).[8] Across all of these groupings, what stands out is a common thread of power imbalance; that is, between those who collect and hold data, and those whom the data is ostensibly about. The new forms of categorisation enabled by the collection of big data are often created without our knowledge and are "based on criteria that do not necessarily correspond to lived experience" (Dencik, Hintz, Redden, & Warne, 2017, p. 734). Ultimately, a truly ethical approach to big data needs to move beyond mere analysis of what particular algorithms achieve—via the collection and manipulation of personal information—towards a more complex interrogation of what gets lost amidst the noise.

The obvious answer to this problem is to build greater equity into infrastructure systems. But how? One possible way is to transform the organising principles of data ethics by moving them away from a consideration of the 'average person' and rather highlighting those who are the most vulnerable and marginalised. It can be argued that this motion foregrounds the relevance of *justice* within big data practices by focusing on the deconstruction of power asymmetries and marginalisation (Taylor, 2017, p. 20; Johnson, 2014; Heeks & Renken, 2018). The problem with this approach, however, is that any form of "data-centric rationality" is always tied to the context of its production; it should therefore be understood as "an expression of the coloniality of power" (Ricaurte, 2019, p. 351). Within this regime, data relations can be defined as "new types of human relations that enable the extraction of data for commodification" (Couldry & Mejias, 2019, p. 337). This extraction is achieved in such a way as to over-surveil particular marginalised citizens, expel particular people from the social order and to suppress or eradicate alternative viewpoints and epistemologies (Escobar, 2017; Santos, 2009). Drawing together these and other effects, Ricaurte (2019, p. 351) notes how this trend "has led to new forms of colonization through data, grounded in material infrastructures and symbolic constructions that reinforce these practices." Thus, the task of reorganising the power structures that underpin big data systems—such as data extraction, storage, processing and analysis—requires a much broader and more rigorous process that must be produced, from the outset, through a decolonial lens (Arora, 2019). Another proposal for tackling some of the more pressing ethical concerns around big data is to reverse the dominant narrative by sharpening the focus on 'small data.' As Lupton (2014, p. 4) has pointed out: "while critical data studies often focus[es] on big data, there is also need for critical approaches to 'small' or personal data, the type of information that people collect on themselves." This was the known intention, for example, of a 2015 special issue of GeoJournal, appropriately titled *What's So Big about Big Data?*. The collection took the "end of theory" as its starting point of provocation for analysing the "epistemic limits of Big Data and accentuating the emerging social, political, and analytic challenges posed by Big Data research and analysis" (Burns & Thatcher, 2014, p. 446).

CONCLUSION

Despite the intensifying critical attention that big data's privacy politics are attracting, it is important to remind ourselves that the contours of big data have changed in the past and will continue to change, particularly with regards to shifting definitions of what privacy means both at an individual and a societal level. Ideas around what privacy means in relation to data must be continually reconsidered via the lens of questions of access, equity, ethics and accountability. Most importantly, this process needs to foreground the assumed 'bigness' of big data by challenging the notion that 'bigger' is necessarily better when it comes to research and knowledge production. A useful way out of this dilemma is to provide ethical and meaningful frameworks that navigate the complex connections between data and social phenomena. Embracing this will not only work to foreground the now generally-accepted notion that data is not ever merely 'raw' material, as well as assisting to build more ethical data infrastructures and analytical methods into our day-to-day practices as researchers, teachers, practitioners and consumers.

8. The Data Justice Lab does, however, acknowledge that in some cases the data harm examples listed could fit into several categories simultaneously (Redden & Brand, 2018).

REFERENCES:

Andrejevic, M. (2014). Big Data, Big Questions: Big Data Divide. *International Journal of Communication*, (8), 1673-1689.

Arora, P. (2019). Decolonizing Privacy Studies. *Television & New Media*, 20(4), 366-378. doi.org/10.1177/1527476418806092

Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. London, UK: The MIT Press.

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

Burns, R., & Thatcher, J. (2014). Guest Editorial: What's so big about Big Data?: Finding the spaces and perils of Big Data. *GeoJournal*, 80(4), 445-448.

Cheney-Lippold, J. (2017). *We Are Data: Algorithms and the Making of Our Digital Selves.* New York: New York University Press.

Couldry, N., & Mejias, U. A. (2019). Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *Television & New Media*, 20(4), 336-349. doi.org/10.1177/1527476418796632

Dencik, L., Hintz, A., Redden, J. and Warne, H. (2018) Data Scores as Governance: Investigating uses of citizen scoring in public services. Retrieved from https://datajustice.files.wordpress.com/2018/12/data-scores-as-governance-project-report2.pdf

Doreski, W. (1999). *Robert Lowell's Shifting Colors: The Poetics of the Public and the Personal*. Athens, USA: Ohio University Press.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. London, UK: St Martin's Press.

Escobar, A. (2017). *Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds*. North Carolina, USA: Duke University Press.

Ferguson. AG. (2015). Big Data and Predictive Reasonable Suspicion. *University of Pennsylvania Law Review*, 163(2): 327-410.

Ferguson. A. G. (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. New York, USA: NYU Press.

Foucault, M. (1975). *Discipline and Punish: The Birth of the Prison*. New York, USA: Pantheon Books.

Gangadharan, S. P. (2012). Digital inclusion and data profiling. *First Monday*, 17(5).

Gellert, R., & Gutwirth, S. (2013). The legal construction of privacy and data protection. *Computer Law and Security Review*, 29(5): 522-530.

Goffman, E. (1969). Strategic Interaction. Philadelphia, USA: University of Pennsylvania Press.

González Fuster, G. (2014). *The Emergence of Personal Data Protection as a Fundamental Right of the EU*. Heidelberg, Germany: Springer.

Harford, T. (2014). Big data: A big mistake? *Significance: Royal Statistical Society*, 11(5): 14-19. doi.org/10.1111/j.1740-9713.2014.00778.x

Heeks, R., & Renken, J. (2018). Data justice for development: What would it mean?. *Information Development*, 34(1), 90-102.

Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2017) Digital Citizenship and Surveillance — Introduction. *International Journal of Communication*, 11, 731-739.

Ioannidis, J. P. A. (2013). Informed Consent, Big Data, and the Oxymoron of Research That is Not Research. *The American Journal of Bioethics*, 13(4), 40-42.

Johnson, J. A. (2014). From Open Data to Information Justice. *Ethics and Information Technology*, 16(4), 263-274.

Koops, B-J. (2014). The trouble with European data protection law. *International Data Privacy Law,* 4(4), 250-261.

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks. *Proceedings of the National Academy of Sciences of the United States of America*, 111(24), 8788-8790.

Kulhari, S. (2018). *Building-Blocks of a Data Protection Revolution: The Uneasy Case for Blockchain Technology to Secure Privacy and Identity.* Baden-Baden, Germany: Nomos.

Lupton, D. (2014). You Are Your Data: Self-Tracking Practices and Concepts of Data. Retrieved from http://papers.ssrn.com/sol3/Papers. cfm?abstract_id=2534211.

Lyon, D. (2003). Surveillance as Social Sorting: Computer Codes and Mobile Bodies. In D. Lyon (Ed.), *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination* (pp. 13-30). London, UK and New York, USA: Routledge.

Mann, M., & Daly, A. (2018). (Big) Data and the North-in-South: Australia's Informational Imperialism and Digital Colonialism. *Television & New Media*, 20(4), 379-395. doi.org/10.1177/1527476418806091

Matzner, T. (2014). Why privacy is not enough privacy in the context of "ubiquitous computing" and "big data". *Journal of Information, Communication and Ethics in Society*, 12(2), 93-106.

Mayer-Schönberger, V., & Kenneth, C. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. New York, USA: Houghton Mifflin Harcourt.

Mosco, V. (2015). *To the Cloud: Big Data in a Turbulent World*. Boulder, UK: Paradigm.

Mosco, V. (2016). Marx in the Cloud. In V. Mosco (Ed.), *Marx in the Age of Digital Capitalism*. Leiden, Nederland: Brill.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, USA: NYU Press.

Parsons, C. (2015). Beyond Privacy: Articulating the Broader Harms of Pervasive Mass Surveillance. *Media and Communication*, 3(3), 1-11. doi.org/10.17645/mac.v3i3.263

Powles, J., & Nissenbaum, H. (2018). The Seductive Diversion of 'Solving' Bias in Artificial Intelligence. Retrieved from https://medium.com/s/story/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53

PrivazyPlan. (2018). Article 6 EU GDPR "Lawfulness of processing." Retrieved from http://www.privacy-regulation.eu/en/article-6-lawfulness-of-processing-GDPR.htm

Redden, J. (2018). The Harm That Data Do. Retrieved from https://www.scientificamerican.com/article/the-harm-that-data-do/

Redden, J., & Brand, J. (2018). Data Harm Record. Retrieved from https://datajusticelab.org/data-harm-record/

Ricaurte, P. (2019). Data Epistemologies, The Coloniality of Power, and Resistance. *Television & New Media*, 20(4), 350-365. doi.org/10.1177/1527476419831640

Richterich, A. (2018). *The Big Data Agenda: Data Ethics and Critical Data Studies.* London, UK: University of Westminster Press.

Rothstein, M., & Shoben, A. (2013). Does Consent Bias Research? T*he American Journal of Bioethics*, 13(4), 27-37.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2016). Automation, Algorithms, and Politics | When the Algorithm Itself is a Racist: Diagnosing Ethical Harm in the Basic Components of Software. *International Journal of Communicatio*n, 10, 4972-4990.

Santos, B. (2009). *An epistemology of the South: the reinvention of knowledge and social emancipation*. Mexico City, Mexico: Siglo XXI.

Schroeder, R. (2018). Social Theory after the Internet: Media, Technology, and Globalization. London, UK: UCL Press.

Schroepfer, M. (2018). An Update on Our Plans to Restrict Data Access on Facebook [Web log post]. Retrieved from https://newsroom.fb.com/news/2018/04/restricting-data-access/

Taylor, L. (2017). What Is Data Justice? The Case for Connecting Digital Rights and Freedoms Globally. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2918779

Tene, O., & Polonetsky, J. (2012). Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property*, 239, 243-251.

van der Ploeg, I. (2007). Genetics, biometrics and the informatization of the body. *Annali dell'Istituto Superiore di Sanita*, 43(1), 44-50.

Warren, S., & Brandeis, L. (1890). The Right to Privacy. *Harvard Law Review*, 193(4), 193-220.

Westin, A. F. (1967). Privacy and Freedom. *Washington and Lee Law Review*, 25(1), 7.