

# Critical data studies: An introduction

Big Data & Society  
 July–December 2016: 1–7  
 © The Author(s) 2016  
 DOI: 10.1177/2053951716674238  
 bds.sagepub.com



Andrew Iliadis<sup>1</sup> and Federica Russo<sup>2</sup>

## Abstract

Critical Data Studies (CDS) explore the unique cultural, ethical, and critical challenges posed by Big Data. Rather than treat Big Data as only scientifically empirical and therefore largely neutral phenomena, CDS advocates the view that Big Data should be seen as always-already constituted within wider data assemblages. Assemblages is a concept that helps capture the multitude of ways that already-composed data structures inflect and interact with society, its organization and functioning, and the resulting impact on individuals' daily lives. CDS questions the many assumptions about Big Data that permeate contemporary literature on information and society by locating instances where Big Data may be naively taken to denote objective and transparent informational entities. In this introduction to the *Big Data & Society* CDS special theme, we briefly describe CDS work, its orientations, and principles.

## Keywords

Critical Data Studies, Big Data, data science, data ethics, data subjects

## Introduction

Data are a form of power. Organizations own vast quantities of user information and hold lucrative data capital (Yousif, 2015), wield algorithms and data processing tools with the ability to influence emotions and culture (Gillespie, 2014; Kramer et al., 2016; Striphos, 2015), and researchers invoke data in the name of scientific objectivity while often ignoring that data are never raw but always “cooked” (Gitelman, 2013). There is evidence that data are surreptitiously extracted from data subjects (Hauge et al., 2016; Metcalf and Crawford, 2016), hijacked to serve agendas that benefit research and industry (Ioannidis, 2005, 2016), and compromised by the interests of not only powerful business organizations but also hackers and rogue agents (Coleman, 2014; Elmer et al., 2015). While data are all of the above and more, they are also conspicuous in their absence—a lack of data is another indication of power, the power not to look or to remain hidden (Brunton and Nissenbaum, 2015; Flyverbom et al., 2016). In their presence and absence, data are always-already active and never neutral, part of an information geography (Graham, 2014, 2015) that is always in flux.

Current research trends in the social and natural sciences indicate a general prioritization of data-intensive and positivistic approaches over long-held

postpositivist and critical approaches (Kitchin, 2015). Discourses and practices surrounding the Big Data revolution (Mayer-Schönberger and Cukier, 2013) feature an emerging variety of new and inventive data science techniques that seek to further scientific inquiry by collecting large amounts of data, directing researchers to novel observations and findings. Some arguments in favor of data-intensive studies situate Big Data science as capable of overthrowing theory (Anderson, 2008) and providing fine-grained analyses that no longer require the critical eye of postpositivist thinking. Yet, as some have noted, “Big Data” remains a metaphor for a set of practices (Puschmann and Burgess, 2014) that are in need of a critical ethos to problematize inherent assumptions about data that pervade current discourses in the natural and social sciences. Big Data are connected to the world in a variety of contexts that exist “beyond” the realm of traditional data science.

<sup>1</sup>University of Ontario Institute of Technology, Canada

<sup>2</sup>University of Amsterdam, The Netherlands

### Corresponding author:

Andrew Iliadis, University of Ontario Institute of Technology, 2000 Simcoe St. N., Oshawa, ON L1H 7K4, Canada.

Email: andrew.iliadis@uoit.ca



Big Data must remain open to cultural, ethical, and critical perspectives, particularly when viewed as a modern archive of data facts and data fictions. Data, along with its sciences and infrastructures, are informed by specific histories, ideologies, and philosophies that tend to remain hidden, though there have been recent calls for inquiry into these domains (Beer, 2016; boyd and Crawford, 2012; Crawford et al., 2014; Floridi, 2011; Kitchin, 2014, 2015). Further, issues of causality (Illari and Russo, 2014), quality (Floridi and Illari, 2014), security (Taddeo, 2013; Taddeo and Floridi, 2015), and uncertainty (Leonelli, 2015) continue to provoke debate among Big Data researchers, practitioners, and their critics. As the product of multiple sites of work, layered analytic techniques, experimental practices, and various competing discourses, Big Data are susceptible to losing provenance and their ability to be “about” only one thing, their origins and interpretations becoming multiple and conflicting as metadata are mixed with primary, secondary, and derivative data. Such a confluence of data sources and meanings inevitably leads to data disorder, the potential for harm to data subjects, and the need for strong ethical investigations into data and its discontents. Big Data belong to a web of subjects, institutions, texts, and authors that tend to remain invisible to researchers who prefer to treat Big Data science as a new form of positivism—but the “data” of “Big Data” are not always the whole story. As Foucault famously put it in *The Archaeology of Knowledge*, the figures that populate a field do not communicate only by the logical successions of propositions but also by the “positivity of their discourse” which defines a field where “formal identities, thematic continuities, translations of concepts, and polemical interchanges may be deployed” (2002: 143). Similarly, Big Data must be challenged by acknowledging the limitations of the positivity of its discourse and the realities of the shifting information infrastructures (Bowker et al., 2010), multiple data subjects and their rights (Haraway, 1991; Jones, 2016), deep information histories (Beniger, 1989), work and power (Zuboff, 1988), and hybrid digital cultures (Striphas, 2016) that underpin it.

### Critical data studies (CDS)

One way that a critical approach to Big Data contributes to knowledge is by helping define the questions that inform epistemological frameworks around social issues related to data. A critical approach to Big Data investigates meta-theoretical modes of conversation and styles of scientific thinking (Hacking, 1994) that pervade data science—it does not contribute to knowledge in the positivistic sense but instead analyzes the ground upon which positivistic Big Data science stands.

How do Big Data inflect and interact with society, social processes, and how we come to measure and interact with them?

The nascent field of CDS is a formal attempt at naming the types of research that interrogate all forms of potentially depoliticized data science and to track the ways in which data are generated, curated, and how they permeate and exert power on all manner of forms of life. In what is already a classic text, boyd and Crawford (2012) proposed a key set of critical questions for Big Data. Going further, Crawford et al. (2014) edited a special collection that built on those original questions by provoking new inquiries into Big Data critique, including issues related to politics, ethics, and epistemology. Dalton and Thatcher (2014) made the original call for CDS and provided the first explicit reference to the field by asking “what does a critical data studies look like?” Kitchin and Lauriault (2014) offered an answer to Dalton and Thatcher’s question and proposed that CDS should study “data assemblages,” that is “the technological, political, social and economic apparatuses and elements that constitutes and frames the generation, circulation and deployment of data” (1). Before and after those publications, CDS has covered a wide area of communications inquiry, including data power issues in social media, apps, the Internet, web, and platforms, but also and equally importantly statistics, policy, research, and organization. In every way that data are organized in a communicative context, CDS—as a clear call for the critical investigation of Big Data science—has coalesced around researchers ready to deploy pronounced critical frameworks in order to foreground data’s power structures. Of course, such a field understandably runs the risk of being overly broad and presumptuously inclusive. As Dalton et al. (2016) note, CDS might offend researchers who point out that all forms of research are critical and create a false separation between critical theory and data science. As such, CDS continues to remain an inclusive field that is open to self-critique and dialog, itself politicized in its quest to politicize Big Data. At the very least, the amorphous groups of individuals, texts, projects, and institutions that seek a specific and pronounced critical engagement with Big Data science now have a name to use.

The multidimensionality of possible critiques of Big Data science grows out of the plurality of data themselves. In their ability to provide interpretations of reality, data are apprehended through various levels of informational abstraction (Floridi, 2011) that frame what data are about. Such frameworks and perspectives on Big Data are multiple and diverse and may attend to any number of apparatuses that reflect specific subject positions. Levels of informational abstraction—the

product of positionalities that constrain and afford what data can be about—are a gateway into the multiple roles that data play and the ways that abstraction may be adopted, manipulated, or repurposed for any number of aims. Choosing a level of abstraction from which to view Big Data alters the types of conversations that can be had about data, its aims, and functions.

As noted by Kitchin (2014) and Kitchin and Lauriault (2014), the subjects of CDS are the sociotechnical “data assemblages” that make up Big Data. The apparatus and elements of a data assemblage may include systems of thought, forms of knowledge, finance, political economy, governmentalities and legalities, materialities and infrastructures, practices, organizations and institutions, subjectivities and communities, places, and the marketplace where data are constituted. Assemblages should be understood as structures that emerge as constitutive of Big Data, viewed from a variety of social positions at multiple scales (local, national, international) that exert power. Data assemblages are the powerful complex of entities that form the underlying production of Big Data science at multiple levels of abstraction and in a plurality of domains.

### CDS work

So far, CDS has emerged as a loose knit group of frameworks, proposals, questions, and manifestos—something to be expected of fields still in their infancy. What need to be established are long-term projects that take up specific challenges in CDS by proposing critical investigations into Big Data assemblages. In this *Big Data & Society* CDS special theme, we collect a group of articles that seek to build on the earlier work of CDS researchers to expose Big Data science problems in social contexts. These articles and commentaries deal with a variety of themes and issues related to Big Data science, ranging from food and health to policing and the environment. The articles feature many issues and subjects that have been of concern to CDS researchers and offer a glimpse into the types of scholarship that CDS work might continue to produce with renewed attention.

For example, Levy and Johns (2016) note that Big Data can be counterintuitively “weaponized” under the veil of openness and transparency and responsible data practices. While Levy and Johns generally agree with data safety practices, they argue that “legislative efforts that invoke the language of data transparency can sometimes function as ‘Trojan Horses’ through which other political goals are pursued” (1). Through an investigation of the “sound science” initiatives of the 1990s and current efforts to open environmental data to public inspection, they find that “[r]ules that exist

mainly to impede science-based policy processes weaponize the concept of data transparency” (1). Similarly, work in CDS should be attuned to the different ways that data can be hijacked and/or weaponized to substantiate pseudoscientific claims that belie political motivations.

Bronson and Knezevic (2016) take up the issue of Big Data in food and agriculture. They review data applications in the agricultural food sector and note that such analytics tools have “implications for relationships of power between players in the food system (e.g., between farmers and large corporations)” (1), looking at issues such as data ownership in the context of applications like Monsanto’s Weed ID app, as well as the privacy implications of John Deere’s precision agricultural equipment. CDS works in food and agriculture platforms tend to be less visible compared to their more traditional social media counterparts. As such, CDS calls for the critical investigation of data-intensive fields that exist outside the ken of traditional “media theory” literature, such as food and agriculture processing data. While traditional social media platforms such as Facebook and Twitter must remain open to CDS work, CDS must further attend to critical data problems in multiple data science domains, from data science’s use in food and agriculture to Big Data techniques in environmental and financial regulation.

Dalton et al. (2016) offer a welcome and open dialog on data, time, and space. Their contribution takes the form of a three-way interview (a valuable format that should be used more often). CDS originated in the field of geography and this article builds on Dalton and Thatcher’s (2014) original call for CDS work that focuses on problems deeply connected to locality and identity. Issues related to the Big Data divide, data discourses, data subjects, and data corporations are framed as central to CDS. Dalton et al. discuss “the stakes, ideas, responsibilities, and possibilities of critical data studies” (1) and in doing so continue the practice of open, sensitive, and politicized dialog among CDS researchers. As a plural and multifaceted field of inquiry, CDS should continue to be open to such forms of dialog, self-critique, and coinvestigation.

Moving from self-critique to a critical engagement of governmental data practices, Rieder and Simon (2016) discuss data’s influence on truth and objectivity in the science of governance. They highlight a growing interest in evidenced-based policy-making and provide an account of data-driven forms of governance. Should numerical evidence produced by Big Data science serve as mandate for the production of policy and new forms of governance? Such questions are beginning to be addressed among CDS researchers who look to interrogate the ways Big Data are used to support

changes in governmentality and social organization, as well as issues related to social policy and practices.

Another policy issue that should be of interest to CDS scholars is the use of human subjects in research. Metcalf and Crawford (2016) address the fundamentally important question of “Where are human subjects in Big Data research?” In discussing the emerging ethics divide, Metcalf and Crawford chart what they view as the “growing discontinuities between the research practices of data science and established tools of research ethics regulation” (1). Making the claim that certain features of ethics regulations cannot be adequately transferred from biomedical research to data science research, Metcalf and Crawford find that this has led some data science researchers to eschew ethical considerations relating to data subjects. Their article discusses current debates around the USA’s Common Rule regarding the regulation of human subjects research and investigates the regulation of social science research, arguing that “data science should be understood as continuous with social sciences in this regard” (1). In emphasizing the ethical dimensions of public datasets and their subjects, Metcalf and Crawford call attention to a growing problem in Big Data science.

Moving from data subjects to data places, Perng et al. (2016) take up the question of locative media and data-driven computing experiments. They note the various ways in which “exploratory data-driven computing experiments” that use geocoding “seek to find ways to extract value and insight” (1) and raise the concern that such practices often begin from data rather than from theory. They argue that locative media data and computing experiments attempt to derive possible futures while having unintended consequences. They further argue that “using computing experiments to imagine potential urban futures produces effects that often have little to do with creating new urban practices” (1). Rather, Perng et al. note that such experiments serve to promote Big Data science and the notion that data may be repurposed.

Tackling another side of Big Data science and its relationship to different localities, Mulder et al. (2016) look into the growing issue of crowdsourced crisis data and humanitarian work. Their aim is to investigate whether Big Data can contribute to an inclusive humanitarian response during large crises. They argue that Big Data are “socially constructed artefacts that reflect the contexts and processes of their creation” (1) across local and international contexts and analyze Big Data-making processes in the context of the 2010 Haiti and 2015 Nepal earthquakes. They find that “locally based, affected people [...] are marginalized in their ability to benefit from Big Data in support of their own means” (1). As such, their work adds to debates

surrounding the use of Big Data in humanitarian contexts.

Beyond humanitarian social data problems, sociotechnical systems that populate the worlds of economics, finance, and the stock market pose a significant challenge to CDS due to their closed, inaccessible nature. Further, semiautomated systems like the stock market and high frequency trading pose new questions in terms of data subjects and subjectivity. Christiaens (2016) provides a critical inquiry into digital subjectivation in the world of finance, writing that “traders have been steadily integrated into computerized data assemblages, which calls for an ontology that eliminates the distinction between human sovereign subjects and non-human instrumental objects” (1). Building on the work of Maurizio Lazzarato, Christiaens provides a critical take on human-machine interaction, arguing that the high-speed data-driven nature of financial markets subjectivize traders in preconscious ways due to their inability to keep pace with automated transactions. Christiaens argues that CDS must consider processes of digital subjectivation and subjugation that occur when Big Data science is applied to socio-technical systems that are governed by humans and machines.

The theme of subjectivity is raised throughout these papers in part due to the lack of discussion around human subjects in Big Data research. Perhaps the most vulnerable, minority and lower socioeconomic status subjects are affected by Big Data science in often invisible and unforeseen ways. Currie et al. (2016) provide an example of such a case in their analysis of four datasets containing police officer-involved homicide statistics in Los Angeles. Their paper frames “police officer-involved homicide data as a rhetorical tool that can reify certain assumptions about the world and extend regimes of power” (1). Civic data, they argue, can be incorporated into creative community practice and events as a form of dataactivism. Comparing local, regional, and national datasets on police officer-involved homicides in Los Angeles, the authors provide “accounts of the semantics, granularity, scale and transparency” of the data before describing a “counter data action” (1) event held with community members.

Whether subjects can trust Big Data is a reoccurring concern and Symons and Alvarado (2016) take up this question. Applying philosophy of science to software, Symons and Alvarado address some of the epistemological challenges posed by Big Data while addressing the topics of computational modeling and simulation. The authors take up the issue of “epistemic opacity” while investigating the problem of error management and error detection. Paying special attention to the relationship between error and path complexity in



software, the article provides an overview of statistical methods and reviews their limitations.

Finally, the CDS special theme concludes with two articles that critically examine the use of data derived from computational modeling for epidemiology and the study of environmental pollution. Canali (2016) addresses the complicated issue of data-driven science's limitations and connection to causality. He focuses on Big Data and causal knowledge by examining EXPOsOMICS, a European Commission-funded project aiming to improve understanding of the relation between exposure and disease. Canali shows how causal knowledge is necessary for EXPOsOMICS and argues that "data-driven claims about causality are fundamentally flawed" (1), suggesting that causal knowledge must remain a necessary part of Big Data science. Thoreau (2016) examines the use of computational models and their data to determine environmental toxicity and assist in the regulation of chemicals. They conclude that quantitative structure-activity relationship models as causal explanation should be reconsidered by regulators.

## Orientations and principles

Each of the articles in this *Big Data & Society* special theme share core concerns that we view as important to CDS. We will end by summarizing three general orientations and principles.

In his *Nicomachean Ethics*, Aristotle famously refers to an "education for the common good"—a perspective that can nurture care by encouraging a shared understanding of specialized knowledge while emphasizing the importance of collective learning and interaction. The notion of education for the common good deeply informs CDS frameworks which should be built to integrate participatory learning and research. In our view, CDS follows three basic principles derived from this broadly Aristotelean approach: the identification of social data problems, the design of critical frameworks for addressing social data problems, and the application of social solutions to increase data literacy. These three simple principles allow for a collective learning experience where critical approaches can be put to use in specific contexts. CDS should strongly emphasize an applied and participatory approach to learning and view interaction as an important part of the applied learning process. Whether through theory, case studies, field work, or using tools, CDS should have the opportunity to practice representing and intervening with reality (Hacking, 1983).

Beginning with the identification of common social data problems is important in that it grounds data problems in terms of relatable and shared dilemmas—for example, research on problems such as the

derivative nature of online metadata in terms of metadata's ability to potentially identify human users. The identification of social data problems should pair Big Data science with common problems, allowing researchers to consider the shared nature of a problematic and to formulate ways in which it might be commonly articulated. This is not a transparent process and researchers should give ample thought to articulating problematic scenarios involving social data. Critical framework designs include viewing data as interpretive and rhetorical assemblages in the construction of science, institutions, and citizens. Established critical frameworks in CDS such as those oriented around data assemblages are just some of the possible directions for CDS frameworks, though it should not be forgotten that CDS also consist of forms of dataactivism and should contribute to data literacy and data justice. The application of social solutions to increase data literacy and justice involves effecting change by conducting research and sharing that research and the activities that might grow out of it with the public. Importantly, CDS should provide individuals with the necessary tools for becoming more informed and the ability to organize efforts around data justice issues. By maintaining these orientations and principles, CDS should encourage us to think about Big Data science in terms of the common good and social contexts.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Notes

1. This Big Data & Society special theme on CDSs grew out of the Society for the Philosophy of Information's Seventh Workshop, "Conceptual Challenges of Data in Science and Technology" (2015, University College London). <http://www.socphilinfo.org/>

## References

- Anderson C (2008) The end of theory: The data deluge makes the scientific method obsolete. *Wired*, 23 June 2008. Available at: [http://www.wired.com/science/discoveries/magazine/16-07/pb\\_theory](http://www.wired.com/science/discoveries/magazine/16-07/pb_theory) (accessed 9 September 2016).
- Beer D (2016) How should we do the history of Big Data? *Big Data & Society* 3(1): 1–10.
- Beniger J (1989) *The Control Revolution: Technological and Economic Origins of the Information Society*. Cambridge, MA: Harvard University Press.

- Bowker GC, Baker K, Millerand F, et al. (2010) Toward information infrastructure studies: Ways of knowing in a networked environment. In: Hunsinger J, Klastrup L and Allen M (eds) *International Handbook of Internet Research*. Dordrecht: Springer, pp. 97–117.
- boyd D and Crawford K (2012) Critical questions for Big Data: Provocations for a cultural, technological and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.
- Bronson K and Knezevic I (2016) Big Data in food and agriculture. *Big Data & Society* 3(1): 1–5.
- Brunton F and Nissenbaum N (2015) *Obfuscation: A User's Guide for Privacy and Protest*. Cambridge: MIT Press.
- Canali S (2016) Big Data, epistemology and causality: Knowledge in and knowledge out in EXPOsOMICS. *Big Data & Society* 3(2): 1–11.
- Christiaens (2016) Digital subjectivation and financial markets: Criticizing social studies of finance with Lazzarato. *Big Data & Society* 3(2): 1–15.
- Coleman G (2014) *Hacker, Hoaxer, Whistleblower, Spy: The Many Faces of Anonymous*. London: Verso.
- Crawford K, Miltner K and Gray ML (2014) Critiquing Big Data: Politics, ethics, epistemology. *International Journal of Communication* 8: 1663–1672.
- Currie M, Paris BS, Pasquetto I, et al. (2016) The conundrum of police officer-involved homicides: Counter-data in Los Angeles County. *Big Data & Society* 3(2): 1–14.
- Dalton, Taylor and Thatcher (2016) Critical data studies: A dialog on data and space. *Big Data & Society* 3(1): 1–9.
- Dalton C and Thatcher J (2014) What does a critical data studies look like, and why do we care? Seven points for a critical approach to 'Big Data.' *Society & Space Open Site*. Available at: <http://societyandspace.com/material/commentaries/craig-dalton-and-jim-thatcher-what-does-a-critical-data-studies-look-like-and-why-do-we-care-seven-points-for-a-critical-approach-to-big-data/> (accessed 6 September 2016).
- Elmer G, Langlois G and Redden J (2015) *Compromised Data: From Social Media to Big Data*. London: Bloomsbury.
- Floridi L (2011) *The Philosophy of Information*. Oxford: Oxford University Press.
- Floridi L and Illari P (2014) *The Philosophy of Information Quality*. Cham: Springer.
- Flyverbom M, Leonardi P, Stohl C, et al. (2016) The management of visibilities in the digital age—Introduction. *International Journal of Communication* 10(2016): 98–109.
- Foucault M (2002) *The Archeology of Knowledge*. London: Routledge.
- Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge: MIT Press, pp. 167–194.
- Gitelman L (ed.) (2013) *"Raw Data" Is an Oxymoron*. Cambridge: MIT Press.
- Graham M (2014) Internet geographies: Data shadows and digital divisions of labour. In: Graham M and Dutton WH (eds) *Society and the Internet: How Networks of Information and Communication Are Changing Our Lives*. Oxford: Oxford University Press, pp. 99–116.
- Graham M (2015) Information geographies and geographies of information. *New Geographies* 7: 159–166.
- Hacking I (1983) *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.
- Hacking I (1994) Styles of scientific thinking or reasoning: A new analytical tool for historians and philosophers of the sciences. In: Gavroglu K, Christianidis J and Nicolaidis E (eds) *Trends in the Historiography of Science*. Boston, MA: Springer, pp. 31–48.
- Haraway D (1991) *Simians, Cyborgs and Women: The Reinvention of Nature*. New York, NY: Routledge.
- Hauge M, Stevenson MD, Rossmo DK, et al. (2016) Tagging Banksy: Using geographic profiling to investigate a modern art mystery. *Journal of Spatial Science* 61(1): 185–190.
- Illari P and Russo F (2014) *Causality: Philosophical Theory Meets Scientific Practice*. Oxford: Oxford University Press.
- Ioannidis JPA (2005) Why most published research findings are false. *PLoS Medicine* 2(8): e124.
- Ioannidis JPA (2016) Evidence-based medicine has been hijacked: A report to David Sackett. *Journal of Clinical Epidemiology* 73: 82–86.
- Jones ML (2016) *Ctrl + Z: The Right to Be Forgotten*. New York: New York University Press.
- Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Los Angeles, CA: Sage.
- Kitchin R (2015) Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1(1): 1–12.
- Kitchin R and Lauriault TP (2014) Towards critical data studies: Charting and unpacking data assemblages and their work. *The Programmable City Working Paper 2*. Available at: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2474112](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112) (accessed 6 September 2016).
- Kramer ADI, Guillory JE and Hancock JT (2016) Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences* 111(24): 8788–8790.
- Leonelli (2015) What counts as scientific data? A relational framework. *Philosophy of Science* 82(5): 810–821.
- Levy KEC and Johns DM (2016) When open data is a Trojan Horse: The weaponization of transparency in science and governance. *Big Data & Society* 3(1): 1–6.
- Mayer-Schönberger and Cukier (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, MA: Houghton Mifflin Harcourt.
- Metcalf and Crawford (2016) Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society* 3(1): 1–14.
- Mulder F, Ferguson J, Groenewegen P, et al. (2016) Questioning Big Data: Crowdsourcing crisis data towards an inclusive humanitarian response. *Big Data & Society* 3(2): 1–13.
- Perng S-Y, Kitchin R and Evans L (2016) Locative media and data-driven computing experiments. *Big Data & Society* 3(1): 1–12.
- Puschmann C and Burgess J (2014) Metaphors of Big Data. *International Journal of Communication* 8(2014): 1690–1709.

- Rieder and Simon (2016) Datatrust: Or, the political quest for numerical evidence and the epistemologies of Big Data. *Big Data & Society* 3(1): 1–6.
- Striphas T (2015) Algorithmic culture. *European Journal of Cultural Studies* 18(4–5): 395–412.
- Striphas T (2016) Culture. In: Peters B (ed.) *Digital Keywords: A Vocabulary of Information Society and Culture*. Princeton, NJ: Princeton University Press, pp. 70–80.
- Symons J and Alvarado R (2016) Can we trust Big Data? Applying philosophy of science to software. *Big Data & Society* 3(2): 1–17.
- Taddeo M (2013) Cyber security and individual rights, striking the right balance. *Philosophy & Technology* 26(4): 353–356.
- Taddeo M and Floridi (2015) The debate on the moral responsibilities of online service providers. *Science and Engineering Ethics* (online): 1–29.
- Thoreau F (2016) A mechanistic interpretation, if possible: How does predictive modelling causality affect the regulation of chemicals? *Big Data & Society* 3(2): 1–11.
- Yousif M (2015) The rise of data capital. *IEEE Cloud Computing* 2(2): 4.
- Zuboff S (1988) *In The Age of the Smart Machine: The Future of Work and Power*. New York, NY: Basic Books.

This commentary is a part of special theme on Critical Data Studies. To see a full list of all articles in this special theme, please click here: <http://bds.sagepub.com/content/critical-data-studies>.