Big Data in the Life Sciences: Bibliography, Computer and Physical Sciences (Part 2)

Author(s)

Valerie Racine

Description

A bibliography looking at the ethics of big data use in the computer and physical sciences.

Body

Journal Articles

Booch, Grady. 2014. "The Human and Ethical Aspects of Big Data." *IEEE Software* **31 (1):20-22. doi: 10.1109/MS.2014.16.**
Every line of code represents a moral decision; every bit of data collected, analyzed, and visualized has moral implications. The Web extra at http://youtu.be/iY7mU1mtQ08 is an audio podcast of author Grady Booch reading his On Computing column, in which he discusses how every line of code represents a moral decision and how every bit of data collected, analyzed, and visualized has moral implications.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* **15 (5):662-679. doi:**

The era of Big Data has begun. Computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, sociologists, and other scholars are clamoring for access to the massive quantities of information produced by and about people, things, and their interactions. Diverse groups argue about the potential benefits and costs of analyzing genetic sequences, social media interactions, health records, phone logs, government records, and other digital traces left by people. Significant questions emerge. Will large-scale search data help us create better tools, services, and public goods? Or will it usher in a new wave of privacy incursions and invasive marketing? Will data analytics help us understand online communities and political movements? Or will it be used to track protesters and suppress speech? Will it transform how we study human communication and culture, or narrow the palette of research options and alter what 'research' means? Given the rise of Big Data as a socio-technical phenomenon, we argue that it is necessary to critically interrogate its assumptions and biases. In this article, we offer six provocations to spark conversations about the issues of Big Data: a cultural, techno- logical, and scholarly phenomenon that rests on the interplay of technology, analysis, and mythology that provokes extensive utopian and dystopian rhetoric.

Chen, Jinchuan, Yueguo Chen, Xiaoyong Du, Cuiping Li, Jiaheng Lu, Suyun Zhao, and Xuan Zhou. 2013. "Big data challenge: a data management perspective." *Frontiers of Computer Science* **7 (2):157-164. doi: 10.1007/s11704-013-3903-7.**

This paper reviews big data challenges from a data management respective. In particular, the authors discuss big data diversity, big data reduction, big data integration and cleaning, big data indexing and query, and finally big data analysis and mining. Our survey gives a brief overview about big-data-oriented research and problems.

Ekbia, Hamid, Michael Mattioli, Inna Kouper, G. Arave, Ali Ghazinejad, Timothy Bowman, Venkata Ratandeep Suri, Andrew Tsou, Scott Weingart, and Cassidy R. Sugimoto. 2015. "Big data, bigger dilemmas: A critical review." *Journal of the Association for Information Science & Technology* **66 (8):1523-1545. doi: 10.1002/asi.23294.**

The recent interest in Big Data has generated a broad range of new academic, corporate, and policy practices along with an evolving debate among its proponents, detractors, and skeptics. While the practices draw on a common set of tools,

techniques, and technologies, most contributions to the debate come either from a particular disciplinary perspective or with a focus on a domain-specific issue. A close examination of these contributions reveals a set of common problematics that arise in various guises and in different places. It also demonstrates the need for a critical synthesis of the conceptual and practical dilemmas surrounding Big Data. The purpose of this article is to provide such a synthesis by drawing on relevant writings in the sciences, humanities, policy, and trade literature. In bringing these diverse literatures together, we aim to shed light on the common underlying issues that concern and affect all of these areas. By contextualizing the phenomenon of Big Data within larger socioeconomic developments, we also seek to provide a broader understanding of its drivers, barriers, and challenges. This approach allows us to identify attributes of Big Data that require more attention-autonomy, opacity, generativity, disparity, and futurity-leading to questions and ideas for moving beyond dilemmas.

Gilmore, Rick O. 2016. "From big data to deep insight in developmental science " *WIREs: Cognitive Science* 7 (2):112-126. doi: 10.1002/wcs.1379.
The use of the term 'big data' has grown substantially over the past several decades and is now widespread. In this review, I ask what makes data 'big' and what implications the size, density, or complexity of datasets have for the science of human development. A survey of existing datasets illustrates how existing large, complex, multilevel, and multimeasure data can reveal the complexities of developmental processes. At the same time, significant technical, policy, ethics, transparency, cultural, and conceptual issues associated with the use of big data must be addressed. Most big developmental science data are currently hard to find and cumbersome to access, the field lacks a culture of data sharing, and there is no consensus about who owns or should control research data. But, these barriers are dissolving. Developmental researchers are finding new ways to collect, manage, store, share, and enable others to reuse data. This promises a future in which big data can lead to deeper insights about some of the most profound questions in behavioral science.

Goodman, Elizabeth. "Design and ethics in the era of big data." Interactions 21, no. 3 (2014): 22-24. doi: 10.1145/2598902
Outlines the author's ethical concerns in the field of big data where the pervasive collecting, storage, and selling of personal data, and how this tests the accepted professional ethics for designer.

Johnson, Jeffrey A. 2014. "From open data to information justice. *Ethics and Information Technology.* 16 (4):263-274.
This paper argues for subsuming the question of open data within a larger question of information justice, with the immediate aim being to establish the need for rather than the principles of such a theory. The author shows that there are several problems of justice that emerge as a consequence of opening data to full public accessibility, and are generally a consequence of the failure of the open data movement to understand the constructed nature of data. The author examines the problems of the embedding of social privilege in datasets as the data is constructed, the differential capabilities of data users (especially differences between citizens and ''enterprise'' users), and the norms that data systems impose through their function as disciplinary systems. In each cases he shows that open data has the quite real potential to exacerbate rather than alleviate injustices.

Lazer, D. The rise of the social algorithm. *Science* 348 (6239):1090-1091.doi: 10.1126/science.aab1422
Humanity is in the early stages of the rise of social algorithms: programs that size us up, evaluate what we want, and provide a customized experience. This quiet but epic paradigm shift is fraught with social and policy implications. The evolution of Google exemplifies this shift. It began as a simple deterministic ranking system based on the linkage structure among Web sites—the model of algorithmic Fordism, where any color was fine as long as it was black (1). The current Google is a very different product, personalizing results (2) on the basis of information about past searches and other contextual information, like location. On page 1130 of this issue, Bakshy et al. (3) explore whether such personalized curation on Facebook prevents users from accessing posts presenting conflicting political views.

Liu, Jianzheng, Jie Li, Weifeng Li, and Jiansheng Wu. 2016. "Rethinking big data: A review on the data quality and usage issues." *ISPRS Journal of Photogrammetry & Remote Sensing* 115:134-142. doi: 10.1016/j.isprsjprs.2015.11.006.
The recent explosive publications of big data studies have well documented the rise of big data and its ongoing prevalence. Different types of "big data" have emerged and have greatly enriched spatial information sciences and related fields in terms of breadth and granularity. However, big data brings lots of "big errors" in data quality and data usage, which cannot be used as a substitute for sound research design and solid theories. The authors indicate and summarize the problems faced by current

big data studies with regard to data collection, processing and analysis: inauthentic data collection, information incompleteness and noise of big data, unrepresentativeness, consistency and reliability, and ethical issues. Cases of empirical studies are provided as evidences for each problem. Finally, they propose that big data research should closely follow good scientific practice to provide reliable and scientific "stories", as well as explore and develop techniques and methods to mitigate or rectify those 'big-errors' brought by big data.

Metcalf, Jacob. 2016. "Big Data Analytics and Revision of the Common Rule." ***Communications of the ACM*** **59 (7):31-33. doi: 10.1145/2935882.**
The article discusses possible changes to the Common Rule in research ethics due to the emergence of big data analytics. Topics include a September 2015 Notice of Proposed Rule-Making (NPRM) issued by the U.S. Department of Health and Human Services (HHS), the distinction between practice and research in relation to the 1979 Belmont Report on human subject research, and the ethical category of publicly available datasets containing private data.

Michael, Katina, and Keith W. Miller. 2013. "Big Data: New Opportunities and New Challenges." ***Computer*** **46 (6):22-24. doi: 10.1109/MC.2013.196.**
Introduces some web special material from an issue of Computer *that looks at some of the ethical challenges of big data.*

Michael, Mike, and Deborah Lupton. 2016. "Toward a manifesto for the 'public understanding of big data'." ***Public Understanding of Science*** **25 (1):104-116. doi: 10.1177/0963662515609005.**
This article sketches h a 'manifesto' for the 'public understanding of big data'. On the one hand, this entails such public understanding of science and public engagement with science and technology–tinged questions as follows: How, when and where are people exposed to, or do they engage with, big data? Who are regarded as big data's trustworthy sources, or credible commentators and critics? What are the mechanisms by which big data systems are opened to public scrutiny? On the other hand, big data generate many challenges for public understanding of science and public engagement with science and technology: How do we address publics that are simultaneously the informant, the informed and the information of big data? What counts as understanding of, or engagement with, big data, when big data themselves are multiplying, fluid and recursive? As part of our manifesto, we propose a range of empirical, conceptual and methodological exhortations.

Pankajdeep, Kaur, and Awal Monga. 2015. "Big Data Management." *International Journal of Advance Foundation And Research In Science & Engineering* **1 (Special Issue).**

Big Data Management (BDM) is the governance and management of huge volumes of all types of data. Big data management is the drastic change to technology that will help to make a better society and the industrial sector. This document deals with the different key factors kept in mind while the management of big data so that organisations can get secure and error free accurate data. This document gives the steps which can be used for managing Big Data and also the role of Hadoop in managing it. Different types of data are increasing and realigning so that industries can fully leverage big data, not just manage it. At the same time, big data must evidently find a permanent and an important place in enterprise data management. Big Data Management deals with the integration, manipulation, quality and governance and management of Big Data including the key factors- Volume, Velocity and Variety of Big Data.

Pileggi, Salvatore F. 2016. "Is Big Data the New "God" on Earth? [Opinion]." *IEEE Technology and Society Magazine* **35 (1):18-20. doi: 10.1109/MTS.2015.2494258.**

As technology continues to pervade more and more human activities, what roll will big data play in our lives?

Portmess, Lisa, and Sara Tower. 2015. "Data Barns, Ambient Intelligence and Cloud Computing: The Tacit Epistemology and Linguistic Representation of Big Data." *Ethics and Information Technology* **17 (1):1-9.**

This paper explores the tacit epistemology of the rhetoric and representation of Big Data and suggests a richer account of its ambiguities and the paradox of its real world materiality. We argue that Big Data should be recognized as manifesting multiple and conflicting trajectories that reflect human intentionality and particular patterns of power and authority. Such patterns require attentive exploration and moral appraisal if we are to resist simplistic informationist ontologies of Big Data, and the subtle forms of control in the political ecology of Big Data that undermine its promise as transformational knowledge.

Qiu, Jack Linchuan. 2015. "Reflections on Big Data: 'Just because it is accessible does not make it ethical'." *Media, Culture & Society* **37 (7):1089-1094. doi: 10.1177/0163443715594104.**

Drawing from observations in China and from world history, this is a reflection on boyd and Crawford's provocation on social problems related to Big Data, especially 'Just because it is accessible does not make it ethnical'.

Rozier, K. Y., and E. W. D. Rozier. 2014. "Reproducibility, correctness, and buildability: The three principles for ethical public dissemination of computer science and engineering research." Ethics in Science, Technology and Engineering, 2014 IEEE International Symposium on, 23-24 May 2014.
The authors propose a system of three principles of public dissemination, which we call reproducibility, correctness, and buildability, and make the argument that consideration of these principles is a necessary step when publicly disseminating results in any evidence-based scientific or engineering endeavor. They examine how these principles apply to the release and disclosure of the four elements associated with computer science research: theory, algorithms, code, and data.

Rubel, Alan, and Kyle M. L. Jones. 2016. "Student privacy in learning analytics: An information ethics perspective." *Information Society* 32 (2):143-159. doi: 10.1080/01972243.2016.1130502.
Higher education institutions have started using big data analytics tools. By gathering information about students as they navigate information systems, learning analytics employs techniques to understand student behaviors and to improve instructional, curricular, and support resources and learning environments. However, learning analytics presents important moral and policy issues surrounding student privacy. The authors argue that there are five crucial questions about student privacy that we must address in order to ensure that whatever the laudable goals and gains of learning analytics, they are commensurate with respecting students' privacy and associated rights, including (but not limited to) autonomy interests. They address information access concerns, the intrusive nature of information-gathering practices, whether or not learning analytics is justified given the potential distribution of consequences and benefits, and issues related to student autonomy.

Schroeder, Ralph, and Josh Cowls. 2014. "[Big data, ethics, and the social implications of knowledge production]." *GeoJournal*.

This position paper addresses current debates about data in general, and big data specifically, by examining the ethical issues arising from advances in knowledge production. Typically, ethical issues such as privacy and data protection are discussed in the context of regulatory and policy debates. Here we argue that this

overlooks a larger picture whereby human autonomy is undermined by the growth of scientific knowledge. To make this argument, we first offer definitions of data and big data, and then examine why the uses of data-driven analyses of human behaviour in particular have recently experienced rapid growth. Next, we distinguish between the contexts in which big data research is used, and argue that this research has quite different implications in the context of scientific as opposed to applied research. We conclude by pointing to the fact that big data analyses are both enabled and constrained by the nature of data sources available. Big data research will nevertheless inevitably become more pervasive, and this will require more awareness on the part of data scientists, policymakers and a wider public about its contexts and often unintended consequences.

Sula, Chris Alen. 2015. "Research Ethics in an Age of Big Data." *Bulletin of the Association for Information Science & Technology* **42 (2):17-21. doi: 10.1002/bul2.2016.1720420207.**
The era of big data introduces new considerations into the traditional context of research ethics. Ethical questions may be considered in terms of accuracy, humane treatment, informed participants and the necessity and applicability of the work, but big data complicates these issues. Since social media participants reflect certain demographic features, data drawn from those sources should not be taken to represent the general population. Big data collection may be more invasive than necessary due to easy access, and consent may be nonexistent. Data that was once anonymous may become identifiable, last indefinitely and conflict with goals for publication. Ways to respect ethics in big data research include involving participants throughout the process, avoiding collecting information that should remain private, notifying participants of their inclusion and providing them options to correct or delete personal information, and using public channels to disseminate research.

Swan, Melanie. 2013. "The Quantified Self: Fundamental Disruption in Big Data Science and Biological Discovery." *Big Data* **1 (2):85-99. doi: 10.1089/big.2012.0002.**
A key contemporary trend emerging in big data science is the quantified self (QS)–individuals engaged in the self-tracking of any kind of biological, physical, behavioral, or environmental information as n=1 individuals or in groups. There are opportunities for big data scientists to develop new models to support QS data collection, integration, and analysis, and also to lead in defining open-access

database resources and privacy standards for how personal data is used. In the long-term future, the quantified self may become additionally transformed into the extended exoself as data quantification and self-tracking enable the development of new sense capabilities that are not possible with ordinary senses. The individual body becomes a more knowable, calculable, and administrable object through QS activity, and individuals have an increasingly intimate relationship with data as it mediates the experience of reality.

Tractenberg, Rochelle, Andrew Russell, Gregory Morgan, Kevin FitzGerald, Jeff Collmann, Lee Vinsel, Michael Steinmann, and Lisa Dolling. 2015. "Using Ethical Reasoning to Amplify the Reach and Resonance of Professional Codes of Conduct in Training Big Data Scientists." *Science & Engineering Ethics* **21 (6):1485-1507. doi: 10.1007/s11948-014-9613-1.**
This normative paper describes an approach to conceptualizing ethical reasoning and integrating it into training for Big Data use and research. Our approach is based on a published framework that emphasizes ethical reasoning rather than topical knowledge. We describe the formation of professional community norms from two key disciplines that contribute to the emergent field of Big Data: computer science and statistics. Historical analogies from these professions suggest strategies for introducing trainees and orienting practitioners both to ethical reasoning and to a code of professional conduct itself. We include two semester course syllabi to strengthen our thesis that codes of conduct (including and beyond those we describe) can be harnessed to support the development of ethical reasoning in, and a sense of professional identity among, Big Data practitioners.

Ulltveit-Moe, Nils. 2014. "A Roadmap towards Improving Managed Security Services from a Privacy Perspective." *Ethics and Information Technology* **16 (3):227-240.**
This paper proposes a roadmap for how privacy leakages from outsourced managed security services using intrusion detection systems can be controlled. The paper first analyses the risk of leaking private or confidential information from signature-based intrusion detection systems. It then discusses how the situation can be improved by developing adequate privacy enforcement methods and privacy leakage metrics in order to control and reduce the leakage of private and confidential information over time. Such metrics should allow for quantifying how much information that is leaking, where these information leakages are, as well as showing what these leakages mean. This includes adding enforcement mechanisms ensuring that

operation on sensitive information is transparent and auditable. The data controller or external quality assurance organisations can then verify or certify that the security operation operates in a privacy friendly manner. The roadmap furthermore outlines how privacy-enhanced intrusion detection systems should be implemented by initially providing privacy-enhanced alarm handling and then gradually extending support for privacy enhancing operation to other areas like digital forensics, exchange of threat information and big data analytics based attack detection.

Zwitter, Andrej. 2014. "Big Data ethics." ***Big Data and Society*** **1 (2). doi: 10.1177/2053951714559253.**

The speed of development in Big Data and associated phenomena, such as social media, has surpassed the capacity of the average consumer to understand his or her actions and their knock-on effects. We are moving towards changes in how ethics has to be perceived: away from individual decisions with specific and knowable outcomes, towards actions by many unaware that they may have taken actions with unintended consequences for anyone. Responses will require a rethinking of ethical choices, the lack thereof and how this will guide scientists, governments, and corporate agencies in handling Big Data. This essay elaborates on the ways Big Data impacts on ethical conceptions.

Contributor(s)

Valerie Racine
Karin Ellison
Joseph Herkert

Rights

Use of Materials on the OEC
License
CC BY-NC-SA

Resource Type

Bibliography

Parent Collection

Big Data in the Life Sciences Collection

Topics

Big Data

Controversies

Discipline(s)

Computer Sciences
Computer, Math, and Physical Sciences