Big Data in the Life Sciences: Bibliography (Part I)

Author(s)

Valerie Racine

Year

2017

Description

A bibliography of guidelines, books, reports and journal articles looking at the use and ethical issues associated with big data in computer science, the information sciences, life and environmental sciences, and the social sciences.

Body

Due to the length of this bibliography, the materials on this page include policies, guidelines, websites and books and reports looking at the ethics of big data, and the sub-pages include journal articles divided by field, namely computer science, life and environmental, and the social sciences (this includes the information sciences).

Policies and Guidelines

Dittrich, David and Erin Kennally. *The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research*. **Washington, D.C.: United States Department on Homeland Security.** *This report proposes a framework for ethical guidelines for computer and*

information security research, based on the principles set forth in the 1979 Belmont Report, a seminal guide for ethical research in the biomedical and behavioral sciences.

Metcalf, Jacob, Emily F. Keller and danah boyd. 2016. *Perspectives on Big Data, Ethics and Society*. **Council for Big Data, Ethics and Society.**
This report addresses concrete manifestations of these disjunctions between big data research methods and existing research ethics paradigms. It identifies policy changes that would encourage greater engagement and reflection on ethics topics and indicates a number of pedagogical needs for data science instructors, and endeavored to fulfill some of them. The report also explores cultural and institutional barriers to collaboration between ethicists, social scientists, and data scientists in academia and industry around ethics challenges. Overall, the report's recommendations are geared toward those who are invested in a future for data science, big data analytics, and artificial intelligence guided by ethical considerations along with technical merit.

United States, National Science Research Council, Subcommittee on Networking and Information Technology Research and Development. 2016. ***The Federal Big Data Research and Development Strategic Plan.***
Issued in May of 2016, this plan provides guidance for developing or expanding federal data and research and development plans, It includes a short section on ethics on pgs. 27-28.

Web Sites

Big Data Ethics
Developed as part of a December 2015 conference on the ethics of big data in research, this site has a number of excellent materials including short papers called "firestarters" and longer papers from the conference looking at methods for providing adequate ethical review processes for big data research in corporate, non-profit, and other non-academic settings.

Council for Big Data and Society
In collaboration with the National Science Foundation, the Council for Big Data, Ethics, and Society was started in 2014 to provide critical social and cultural perspectives on big data initiatives. The Council brings together researchers from

diverse disciplines to address issues such as security, privacy, equality, and access in order to help guard against the repetition of known mistakes and inadequate preparation. Through public commentary, events, white papers, and direct engagement with data analytics projects, the Council will develop frameworks to help researchers, practitioners, and the public understand the social, ethical, legal, and policy issues that underpin the big data phenomenon. Includes a good [bibliography](#) *on big data and ethics from members of the council.*

The Brain Mapping Initiatives: Foundational Issues

On December 6-7, New York University will host a conference devoted to foundational issues raised by recent brain mapping initiatives, such as the BRAIN initiative, the Human Brain project, the Human Connectome Project, and the Allen Brain Atlas. What can mapping the brain tell us about the human mind? What are the ethical implications? These issues will be discussed by leading cognitive neuroscientists and philosophers.

Books and Reports

Béranger, Jérôme. 2016. *Ethics in Big Data: the medical datasphere.* London: Elesvier.
Faced with the exponential development of big aata and both its legal and economic repercussions, we are still slightly in the dark concerning the use of digital information. In the perpetual balance between confidentiality and transparency, this data will lead us to call into question how we understand certain paradigms, such as the Hippocratic Oath in medicine. This empirical and ethico-technical approach brings together the first aspects of a moral framework directed toward thought, conscience and the responsibility of citizens concerned by the use of data of a personal nature.

Collmann, Jeff, Soren Adam Matei. 2016. *Ethical Reasoning in Big Data: an exploratory analysis.* Switzerland: Springer.
This book springs from a multidisciplinary, multi-organizational, and multi-sector conversation about the privacy and ethical implications of research in human affairs using big data. The need to cultivate and enlist the public's trust in the abilities of particular scientists and scientific institutions constitutes one of this book's major themes. Dedicated to the practice of ethical reasoning and reflection in action, the book offers a range of observations, lessons learned, reasoning tools, and

suggestions for institutional practice to promote responsible big data research on human affairs.

Davis, Kord and Patterson, Doug. 2012. *Ethics of Big Data.* Farnham, O'Reilly.
This short book examines ethical questions raised by the big data phenomenon, and explains why enterprises need to reconsider business decisions concerning privacy and identity

Lane, Julia, Victoria Stodden, Stefan Bender, and Helen Nissenbaum. 2014. *Privacy, Big Data, and the Public Good: Frameworks for Engagement*: Cambridge University Press.
Massive amounts of data on human beings can now be analyzed. Pragmatic purposes abound, including selling goods and services, winning political campaigns, and identifying possible terrorists. Yet "big data" can also be harnessed to serve the public good: scientists can use big data to do research that improves the lives of human beings, improves government services, and reduces taxpayer costs. In order to achieve this goal, researchers must have access to this data - raising important privacy questions. What are the ethical and legal requirements? What are the rules of engagement? What are the best ways to provide access while also protecting confidentiality? Are there reasonable mechanisms to compensate citizens for privacy loss? The goal of this book is to answer some of these questions. The book's authors paint an intellectual landscape that includes legal, economic, and statistical frameworks. The authors also identify new practical approaches that simultaneously maximize the utility of data access while minimizing information risk.

Mittelstadt , Daniel Brent and Luciano Floridi. 2016. *The Ethics of Biomedical Big Data*. Dordrecht: Springer.
Biomedical big data refers to the analysis of aggregated, very large datasets to improve medical knowledge and clinical care. The book describes the ethical problems posed by aggregation of biomedical datasets and re-use/re-purposing of data, in areas such as privacy, consent, professionalism, power relationships, and ethical governance of big data platforms. Approaches and methods are discussed that can be used to address these problems to achieve the appropriate balance between the social goods of biomedical big data research and the safety and privacy of individuals.

The National Academies of Science, Engineering and Medicine 2010. *Conducting Biosocial Surveys: Collecting, Storing, Accessing, and Protecting Biospecimens and Biodata*. Washington, D.C: National Academies Press.
This report offers findings and recommendations concerning the best approaches in the Biosocial field. The topics covered include: informed consent, privacy issues and the best practice, but also additional legal, ethical, and social issues, as well as practical issues related to the storage, retrieval, and sharing of data.

The National Academies of Science, Engineering and Medicine. 2009. Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. Washington, D.C: National Academies Press.
As digital technologies are expanding the power and reach of research, they are also raising complex issues. These include complications in ensuring the validity of research data; standards that do not keep pace with the high rate of innovation; restrictions on data sharing that reduce the ability of researchers to verify results and build on previous research; and huge increases in the amount of data being generated, creating severe challenges in preserving that data for long-term use.Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age examines the consequences of the changes affecting research data with respect to three issues - integrity, accessibility, and stewardship-and finds a need for a new approach to the design and the management of research projects.

Pimple, Kenneth D. 2014. *Emerging Pervasive Information and Communication Technologies (PICT): Ethical Challenges, Opportunities and Safeguards*. Law, Governance and Technology Series, Vol. 11. Dordrecht: Springer.
This book provides a wide and deep perspective on the ethical issues raised by pervasive information and communication technology (PICT) – small, powerful, and often inexpensive Internet-connected computing devices and systems. It describes complex and unfamiliar technologies and their implications, including the transformative potential of augmented reality, the power of location-linked information, and the uses of "big data," and explains potential threats, including privacy invaded, security violated, and independence compromised, often through widespread and lucrative manipulation.

Rothbern, Marc, Jeramie Scott and Julia Horwitz. 2015. *Privacy in the Modern Age: The Search for Solutions*. New York: The New Press.
This anthology discusses the problems that the use of big data, social media, the

ability to track individuals via phones, and other technologies and new practices pose to individual privacy, and offers some solutions to try and responsibly answer these new questions.

Stodden, Victoria. 2014. *Privacy, Big Data, and the Public Good: Frameworks for Engagement*. New York: Cambridge University Press.
Massive amounts of data on human beings can now be analyzed. Pragmatic purposes abound, including selling goods and services, winning political campaigns, and identifying possible terrorists. Yet "big data" can also be harnessed to serve the public good: scientists can use big data to do research that improves the lives of human beings, improves government services, and reduces taxpayer costs. In order to achieve this goal, researchers must have access to this data - raising important privacy questions. What are the ethical and legal requirements? What are the rules of engagement? What are the best ways to provide access while also protecting confidentiality? Are there reasonable mechanisms to compensate citizens for privacy loss? The goal of this book is to answer some of these questions. The book's authors paint an intellectual landscape that includes legal, economic, and statistical frameworks. The authors also identify new practical approaches that simultaneously maximize the utility of data access while minimizing information risk.

Journal Articles

Ambrose, Meg Leta. 2014. "[Lessons from the Avalanche of Numbers: Big Data in Historical Context](#)." *I/S: A Journal of Law and Policy for the Information Society* 11(2): 201-277.
The big data revolution, like many changes associated with technological advancement, is often compared to the industrial revolution to create a frame of reference for its transformative power, or portrayed as altogether new. This article argues that between the industrial revolution and the digital revolution is a more valuable, yet overlooked period: the probabilistic revolution that began with the avalanche of printed numbers between 1820 and 1840. By comparing the many similarities between big data today and the avalanche of numbers in the 1800s, the article situates big data in the early stages of a prolonged transition to a potentially transformative epistemic revolution, like the probabilistic revolution. The widespread changes in and characteristics of a society flooded by data results in a transitional state that creates unique challenges for policy efforts by disrupting foundational

principles relied upon for data protection. The potential of a widespread, lengthy transition also places the law in a pivotal position to shape and guide big data-based inquiry through to whatever epistemic shift may lie ahead.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data." *Information, Communication & Society* **15 (5):662-679. doi: 10.1080/1369118X.2012.678878.**
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.

Crawford. 2014a. "[The Test We Can—and Should—Run on Facebook](#)." The Atlantic. July 2.
Discusses the Facebook emotional contagion experiment and what it means ethically for social researcher interested in doing large-scale user experimentation using social media and big data.

Crawford, Kate and Jacob Metcalf. 2016. "[Where are Human Subjects in Big Data Research? The Emerging Ethics Divide](#)." *Big Data and Society.*
There are growing discontinuities between the research practices of data science and established tools of research ethics regulation. Some of the core commitments of existing research ethics regulations, such as the distinction between research and practice, cannot be cleanly exported from biomedical research to data science research. Such discontinuities have led some data science practitioners and

researchers to move toward rejecting ethics regulations outright. These shifts occur at the same time as a proposal for major revisions to the Common Rule—the primary regulation governing human-subjects research in the USA—is under consideration for the first time in decades. The authors contextualize these revisions in long-running complaints about regulation of social science research and argue data science should be understood as continuous with social sciences in this regard. The proposed regulations are more flexible and scalable to the methods of non-biomedical research, yet problematically largely exclude data science methods from human-subjects regulation, particularly uses of public datasets.

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.

Fuller, Michael. 2015. "Big Data: new science, new challenges, new digital opportunities." *Zygon: Journal of Religion & Science* **50 (3):569-582. doi: 10.1111/zygo.12187.**
The advent of extremely large data sets, known as 'big data,' has been heralded as

the instantiation of a new science, requiring a new kind of practitioner: the 'data scientist.' This article explores the concept of big data, drawing attention to a number of new issues-not least ethical concerns, and questions surrounding interpretation-which big data sets present. It is observed that the skills required for data scientists are in some respects closer to those traditionally associated with the arts and humanities than to those associated with the natural sciences; and it is urged that big data presents new opportunities for dialogue, especially concerning hermeneutical issues, for theologians and data scientists.

Horvitz, Eric and Deirdre Mulligan. 2015. "Data, privacy, and the greater good." *Science, Policy Forum*, 17 July 2015. 349 (6245): 253-255.
Large-scale aggregate analyses of anonymized data can yield valuable results and insights that address public health challenges and provide new avenues for scientific discovery. These methods can extend our knowledge and provide new tools for enhancing health and wellbeing. However, they raise questions about how to best address potential threats to privacy while reaping benefits for individuals and to society as a whole. The use of machine learning to make leaps across informational and social contexts to infer health conditions and risks from nonmedical data provides representative scenarios for reflections on directions with balancing innovation and regulation.

Ioannidis, John P. A. 2013. "Informed Consent, Big Data, and the Oxymoron of Research That Is Not Research." *American Journal of Bioethics* 13 (4):40-42.
Discusses the issue of how consent bias can distort research results and that in the age of big data, informed consent might or might not be necessary.

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, David. 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.

Martin, Kirsten E. 2015. "[Ethical Issues in the Big Data Industry](#)." *MIS Quarterly Executive* **14 (2):67-85.**
Big Data combines information from diverse sources to create knowledge, make better predictions and tailor services. This article analyzes Big Data as an industry, not a technology, and identifies the ethical issues it faces. These issues arise from reselling consumers' data to the secondary market for Big Data. Remedies for the issues are proposed, with the goal of fostering a sustainable Big Data Industry.

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, 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.

Nissenbaum, Helen. 2015. "Assuring a Role for 'Respect for Context' in Protecting Privacy." In *Privacy in the Modern Age: The Search for Solutions*, edited by Marc Rotenberg, Jeramie Scott, and Julia Horwitz. The New Press.
An excellent introduction to changing concepts of privacy in the field of big data.

Panger, Galen. 2016. "Reassessing the Facebook experiment: critical thinking about the validity of Big Data research." *Information, Communication & Society* 19 (8):1108-1126. doi: 10.1080/1369118X.2015.1093525.
The Facebook experiment of 2014 manipulated the contents of nearly 700,000 users' News Feeds to induce changes in their emotions. This experiment was widely criticized on ethical grounds regarding informed consent. This controversy, however, diverted attention from a more important concern the experiment was intended to address, which is the impact of Facebook use on well-being. In this paper, the author explores the well-being concerns raised by prior research and argue that the experiment does not alleviate them, owing to poor research design. As the question of Facebook's impact on well-being is of great importance, both to Facebook and to society overall, there is a pressing need for more experimental research that is both sensitive to informed consent and carefully designed to yield reliable results. In turn, the lessons of this case have implications for general issues of validity that emerge in Big Data research, now in vogue at major scientific venues.

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.

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.

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.

Voosen, Paul. 2015. "After Facebook Fiasco, Big-Data Researchers Rethink Ethics." *Chronicle of Higher Education* **61 (17):A14-A14.**
The article discusses the impact of controversy over an academic study conducted in collaboration with online social networking company Facebook in which Facebook users' were studied without their knowledge, highlighting concerns about the ethics of big-data research. Topics addressed include the views of professor Jeffrey T. Hancock, who worked with Facebook on the study, as well as criticism of the lack of informed consent in the study.

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
Joseph Herkert
Karin Ellison

Rights

Resource Type

Bibliography

Parent Collection

Big Data in the Life Sciences Collection

Topics

Bias in Research
Big Data
Communicating Science and Engineering
Controversies
Data Management
Privacy and Surveillance

Discipline(s)

Computer, Math, and Physical Sciences
Computer Sciences
Information Sciences
Life and Environmental Sciences