Online Ethics Center
FOR ENGINEERING AND SCIENCE

Big Data Subject Aid

Author(s)

Rachelle Hollander

Year

2016

Description

A short guide to some key resources and readings on the topic of the ethics of big data use.

Body

A straightforward definition of “big data” appears towards the top of a Google search (June 29, 2016) as “extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions.” This is a useful definition; however, the term “data” goes undefined in it. The same source provides a definition of data as “facts and statistics collected together for reference or analysis.” A collection implies a selection; however large, the selected nature of data remains. All data is selected or captured from among myriad possibilities, put together, or constructed.

Data selection is clearly wrong or inadequate when reliant on or inclusive of fabricated or falsified data. There can also be unintended bias in data selection and its selective utilization. Selection of data for analysis or presentation is legitimate only when undertaken on the basis of clear criteria for thinking that in comparison with other related data the selection is less subject to confounding influences or

"noise." One also needs to pay attention to the issues which arise when “cleaning” data for use, or transforming it for use in another environment (a secondary or tertiary use of data collected for a specific purpose).

There is a long history of collecting and using data by governmental and private entities. What’s now of interest is partly the size but really new forms of data analytics which makes data resources of greater use to the collectors but also of greater potential for harms and misuses to human beings and communities and the environment. Data selection and concatenation must be disclosed in reporting the research, to satisfy standards for ethical research practice; but disclosure may not suffice to prevent harms to individuals, groups, or institutions and may indeed have unintended negative consequences.

See topical collection entries “Bias in Research” and “Conflict of Interest.”

Subject Overviews

Pimple, Kenneth D. ed. *Emerging Pervasive Information and Communication Technologies (PICT): Ethical Challenges, Opportunities and Safeguards*. Law, Governance and Technology Series, Vol. 11. Springer (September 2013), Introduction, 1-12.

> This collection provides analysis of the ethical issues posed by pervasive information and communication technologies (PICT). Its perspective is that their development and use should be informed by “anticipatory ethics,” which views technological development as part of socio-technical systems. Concern for the social and ethical consequences of these systems requires attention to the influences of innovations in and on the systems.

Andrejevic, Marc. “Surveillance in the Big Data Era” in *Emerging Pervasive Information and Communication Technologies (PICT): Ethical Challenges, Opportunities and Safeguards*. Law, Governance and Technology Series, Vol. 11. Springer (September 2013), Chapter 4, 55-69.

> Emerging regimes of surveillance and monitoring indicate a change from targeted to generalized surveillance. However, the goals of surveillance remain, although the process now uses an algorithm to isolate or identify the

desired audience. This chapter addresses ethical issues that arise in monitoring populations and the challenge to democracy this monitoring poses.

Johnson, Jeffrey A. 2014. From open data to information justice. *EIT* 16: 263-274.

This paper argues for subsuming the question of open data within a larger question of information justice. Several problems of justice follow from opening data to full public accessibility and the failure of the open data movement to understand the constructed nature of data. Three such problems are the embedding of social privilege in the construction of datasets, the differential capabilities of data users, and the norms of data systems functioning as disciplinary systems. In such cases, open data has the quite real potential to exacerbate rather than alleviate injustices. This necessitates a theory of information justice. There are two complementary directions for such a theory: one defining a set of moral inquiries that can be used to evaluate the justness of data practices, and another exploring the practices and structures that a social movement promoting information justice might pursue.

Lazer, D. The rise of the social algorithm. *Science* 5 June 2015 348:6239. 1090-1091.

Social algorithms are programs intended to provide customized experiences to online users. The question addressed in this article is what are the implications ether this type of personal curating for access to conflicting views and the quality of deliberation in democratic societies.

Crawford, Kate, Mary L. Gray, and Kate Miltner. "Big Data| Critiquing Big Data: Politics, Ethics, Epistemology| Special Section Introduction." *International Journal of Communication* 8 (2014): 1-11.

This introduction asks why big data has gained such remarkable purchase in a range of industries and across academia, at this point in the 21st century. Big data now ranges across a vast terrain that spans health care, astronomy, policing, city planning, and advertising. From the RNA bacteriophages in our bodies to the Kepler Space Telescope, searching for terrorists or predicting cereal preferences, big data is deployed as the term of art to encompass all the techniques used to analyze data at scale. But why has the concept gained such traction now?

Horvitz, E & Mulligan, D. 2015. Data, privacy, and the greater good. *Science, Policy Forum*, 17 July. 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. 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.

Sax, Marijn. 2016. Big Data: Finders keepers, losers weepers? *EIT* 18:1. March, 25-31

Commercial success of big data requires both new technological capabilities and social institutions that allow organizations to "own" these results. This article argues that the ethical assumptions underlying this presumption require justifications. In particular, the author argues that three assumptions are very questionable – that personal data can be separated from its subjects so as to negate any claim to compensation or control; that acquiring the data is legitimate because the transaction has subjects' consent; and finally, if the first two conditions are satisfied, the outcomes must be just.

Policy or Guidance

Crawford, Kate. 2016. A.I.'s white guy problem. *New York Times Sunday Review*. 11.

Many machine learning algorithms incorporate biases that exacerbate inequities. They put particular groups at a disadvantage – ranging from facial recognition problems to biases in assignments of risks of recidivism, even to notification about the availability of high status jobs. Commitments to address these problems from the technical and policy communities are needed.

Dove, Edward S, David Townend, Eric M. Meslin, Martin Bobrow, Katherine Littler, Dianne Nicol, Jantina de Vries, Anne Junker, Chiara Garattini, Jasper

Bovenberg, Mahsa Shabani, Emmanuelle Levesqure, Bartha M. Knoppers. 2016. Ethics Review for International Data-Intensive Research. *Science* 351: 6280, March 25. 1399-1400.

> The authors reviewed a number of approaches to ethics review for large data sets relevant to human subjects and their protection (excluding clinical trials research) and identified three models that could inform a framework allowing mutual recognition of international ethics review. The models are reciprocity, delegation, and federation, and a chart listing advantages and disadvantages and examples of projects for each model is provided.

The National Academies of Sciences, Engineering and Medicine. 2013. *Proposed Revisions to the Common Rule: Perspectives of Social and Behavioral Scientists: Workshop Summary.* Washington, DC: The National Academies Press. http://www.onlineethics.org/Resources/34405.aspx Accessed July 5, 2016.

> The summary focuses on: 1. Evidence on the functioning of the Common Rule and of institutional review boards (IRBs). 2. Types and levels of risk and harms in social and behavioral sciences, and issues of severity and probability of harm. 3. Consent and special populations. 4. Protection of research participants. 5. Multidisciplinary and multisite studies. 6. The purview and roles of IRBs.

Metcalf, Jacob. Computing ethics: Big Data Analytics and Revision of the Common Rule, Communications of the ACM, Vol. 59 No. 7, Pages 31-33, July 2016. http://cacm.acm.org/magazines/2016/7/204018-big-data-analytics-and-revision-of-the-common-rule/fulltext. Accessed 12/13/16

> Summary of issues facing data scientists whose research and practice more and more concerns human beings and human subjects, a domain traditionally thought to concern mainly social and behavioral scientists.

Bibliography

- A useful bibliography on ethical and social aspects arising with big data development, projects, and use, can be found in the Literature section of the website of the Council for Big Data, Ethics, and Society, at

http://bdes.datasociety.net/literature/. While limited to publications that include members of the Council, it contained more than 90 entries when accessed on July 5, 2016.

- For an article examining problems for confidentiality in the era of “big data” see Nate Anderson, “Anonymized data really isn’t – and here’s why not.” Ars Technica 9/8/2009. Accessed 12/13/2016. http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/

Rights

Use of Materials on the OEC

Resource Type

Bibliography

Parent Collection

OEC Subject Aids

Topics

Big Data
Controversies

Discipline(s)

Research Ethics
Information Sciences
Life and Environmental Sciences
Social and Behavioral Sciences
Computer, Math, and Physical Sciences
Authoring Institution
Online Ethics Center