



Online Ethics Center  
FOR ENGINEERING AND SCIENCE

# Big Data in the Life Sciences: Bibliography, Life and Environmental Sciences (Part 3)

## Author(s)

Valerie Racine

## Description

This bibliography includes policies and guidelines, web resources, books, and journal articles looking at issues of big data in the life and environmental sciences.

## Body

### Policies and Guidance

**Rosenblat, Alex. Kate Wikelius, danah boyde, Seeta Peña, Gangadharan, and Corrine Yu. 2014. [Data Health and Civil Rights: Health Primer](#). The Council for Big Data and Society.**

*Developed for a conference held in October of 2014, this is a summary of the role data plays in medicine and insurance, and how people's medical conditions and health habits can be inferred from many sources of data. When inaccurate inferences are made or sensitive personal information is revealed, these inferences can result in individuals not getting fair access to health care and can also erode productive efforts to use data to empower people, personalized medicine, and develop innovations that can advance healthcare.*

## **Books**

**Béranger, Jérôme. 2016. *Ethics in Big Data: the medical datasphere*. London: Elsevier.**

*Faced with the exponential development of big data 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.*

**Goodman, Kenneth W. *Ethics, Medicine and Information Technology: Intelligent machines and the transformation of health care*. New York: Cambridge University Press.**

*Information technology is transforming the practices of medicine, nursing, and biomedical research. The concepts of privacy and confidentiality are evolving as data moves from paper to silicon to clouds. Big data promises financial wealth, as well as riches of information and benefits to science and public health. This book looks at what will become the core curriculum in ethics and health informatics and discusses how ethical issues surrounding data use in this field are likely to surface in research and the workplace both currently and in the next few years.*

**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.

## **Journal Articles**

**Cato, Kenrick D., Walter Bockting, and Elaine Larson. 2016. "Did I Tell You That? Ethical Issues Related to Using Computational Methods to Discover Non-Disclosed Patient Characteristics." *Journal of Empirical Research on Human Research Ethics* 11 (3):214-219. doi: 10.1177/1556264616661611.**

*Widespread availability of electronic health records coupled with sophisticated statistical methods offer great potential for a variety of applications for health and disease surveillance, developing predictive models and advancing decision support for clinicians. However, use of "big data" mining and discovery techniques has also raised ethical issues such as how to balance privacy and autonomy with the wider public benefits of data sharing. Furthermore, electronic data are being increasingly used to identify individual characteristics, which can be useful for clinical prediction and management, but were not previously disclosed to a clinician. This process in computer parlance is called electronic phenotyping, and has a number of ethical implications. Using the Belmont Report's principles of respect for persons, beneficence, and justice as a framework, the authors examine the ethical issues posed by electronic phenotyping.*

**Choudhury, Suparna, Jennifer R. Fishman, Michelle L. McGowan, and Eric T. Juengst. 2014. "Big data, open science and the brain: lessons learned from genomics." *Frontiers in Human Neuroscience* 8:1-10. doi: 10.3389/fnhum.2014.00239.**

*The BRAIN Initiative aims to break new ground in the scale and speed of data collection in neuroscience, requiring tools to handle data in the magnitude of yottabytes (10<sup>24</sup>). The scale, investment and organization of it are being compared to the Human Genome Project (HGP), which has exemplified "big science" for biology. In line with the trend towards Big Data in genomic research, the promise of the BRAIN Initiative, as well as the European Human Brain Project, rests on the possibility to amass vast quantities of data to model the complex interactions between the brain and behavior and inform the diagnosis and prevention of neurological disorders and psychiatric disease. Advocates of this "data driven" paradigm in neuroscience argue that harnessing the large quantities of data generated across laboratories worldwide has numerous methodological, ethical and economic advantages, but it requires the neuroscience community to adopt a culture of data sharing and open access to benefit from them.*

**Chow-white, Peter A., Maggie Macaulay, Anita Charters, and Paulina Chow. 2015. "From the bench to the bedside in the big data age: ethics and**

**practices of consent and privacy for clinical genomics and personalized medicine." *Ethics and Information Technology* 17 (3):189-200. doi: <http://dx.doi.org/10.1007/s10676-015-9373-x>.**

*Scientists and clinicians are starting to translate genomic discoveries from research labs to the clinical setting. In the process, big data genomic technologies are both a risk to individual privacy and a benefit to personalized medicine. There is an opportunity to address the social and ethical demands of various stakeholders and shape the adoption of diagnostic genome technologies. We discuss ethical and practical issues associated with the networking of genomics by comparing how the European Union (EU) and North America understand and practice notions of privacy and consent in research. An overview of international policy suggests the embedding of genomics within digital networks and the Internet creates conditions that challenge the management of privacy and consent in the age of big data. The risks of reidentification, informational harms, and data security vulnerabilities are issues that need to be better addressed in the clinical setting to reconcile the unpredictable pathway of research and practice in the networked information society.*

**Dereli, Türkay, Yavuz Coşkun, Eugene Kolker, Öner Güner, Mehmet Ağırbaşı, and Vural Özdemir. 2014. "Big Data and Ethics Review for Health Systems Research in LMICs: Understanding Risk, Uncertainty and Ignorance—And Catching the Black Swans?" *American Journal of Bioethics* 14 (2):48-50. doi: [10.1080/15265161.2013.868955](https://doi.org/10.1080/15265161.2013.868955).**

*The authors reflects on the ethics review needed for the big data health systems research (HSR), explaining that a proper review needs to separate among the concepts of risk, uncertainty, ignorance and black swan events. Topics include the manner of addressing low- and middle-income countries' (LMICs) HSR and the likelihood of big data governance missing an opportunity for ethics capacity building in LMICs.*

**Dove, E. S., and V. Ozdemir. 2015. "What Role for Law, Human Rights, and Bioethics in an Age of Big Data, Consortia Science, and Consortia Ethics? The Importance of Trustworthiness." *Laws* 4 (3):515-540. doi: [10.3390/laws4030515](https://doi.org/10.3390/laws4030515).**

*The global bioeconomy is generating new paradigm-shifting practices of knowledge co-production, such as collective innovation; large-scale, data-driven global consortia science (Big Science); and consortia ethics (Big Ethics). These bioeconomic*

*and sociotechnical practices can be forces for progressive social change, but they can also raise predicaments at the interface of law, human rights, and bioethics. In this article, we examine one such double-edged practice: the growing, multivariate exploitation of Big Data in the health sector, particularly by the private sector. What role do law, human rights, and bioethics—separate and together—have to play in addressing these predicaments and opportunities in early 21st century science and society? One answer we propose is an intertwined ethico-legal normative construct, namely trustworthiness. By considering trustworthiness as a central pillar at the intersection of law, human rights, and bioethics, we enable others to trust us, which in turn allows different actors (both nonprofit and for-profit) to operate more justly in consortia science and ethics, as well as to access and responsibly use health data for public benefit.*

**Frizzo-Barker, Julie, and Peter A. Chow-White. 2014. "From Patients to Petabytes: Genomic Big Data, Privacy, and Informational Risk." *Canadian Journal of Communication* 39 (4):615-625.**

*Genomic big data is an emerging information technology, which presents new opportunities for medical innovation, as well as new challenges to our current ethical, social, and legal infrastructure. Rapid, affordable whole-genomic sequencing translates patients' most sensitive personal information into petabytes of digital health data. While a biomedical approach traditionally focuses on risks and benefits to the human body, the fields of Communication and Science and Technology Studies (STS) can provide some of the critical and theoretical tools necessary to navigate the newly emerging terrain of the human body as digital code. Core areas of expertise from these fields, including the Internet, the network society, and the social constructions of technology, ground our discussion of the social implications of open access genomic databases, privacy, and informational risk.*

**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.*

**Hall, Melinda C., and Sharona Hoffman. 2016. "[Big Data](#)" in Healthcare has Some Ethicists Concerned." *Medical Ethics Advisor* 32 (6):64-65.**

*A short article discussing some of the privacy and other ethical concerns raised by the Precision Medicine Initiative which sought to recruit 1 million volunteers to share genetic data and biological samples as a way to develop targeted approaches to diseases.*

**Heckmann, Dirk. 2015. "Big Data: Mehr Maschine -- weniger Mensch? Gedanken zu Recht und Ethik einer überspannten Datennutzung." *Jahrbuch fuer Recht und Ethik* 23:17-31.**

**Hoffman, Sharona, and Andy Podgurski. 2013. "Big bad data: law, public health, and biomedical databases." *The Journal of Law, Medicine & Ethics* 41 (1): 56-60.**

*The accelerating adoption of electronic health record (EHR) systems will have far-reaching implications for public health research and surveillance, which in turn could lead to changes in public policy, statutes, and regulations. The public health benefits of EHR use can be significant. However, researchers and analysts who rely on EHR data must proceed with caution and understand the potential limitations of EHRs.*

**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.*

**Hull, Gordon. 2015. "Successful Failure: What Foucault Can Teach Us about Privacy Self-Management in a World of Facebook and Big Data." *Ethics and Information Technology* 17 (2):89-101.**

*The "privacy paradox" refers to the discrepancy between the concern individuals express for their privacy and the apparently low value they actually assign to it when they readily trade personal information for low-value goods online. In this paper, I argue that the privacy paradox masks a more important paradox: the self-management model of privacy embedded in notice-and-consent pages on websites and other, analogous practices can be readily shown to under-protect privacy, even in the economic terms favored by its advocates. The real question, then, is why privacy self-management occupies such a prominent position in privacy law 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.*

**Kaplan, Bonnie. 2015. "Selling Health Data: De-identification, Privacy, and Speech." *Cambridge Quarterly of Healthcare Ethics* 24 (3):256-271.**

*Describes two court cases dealing with the selling of prescription data for pharmaceutical marketing concern privacy from the United, and discusses how these concern privacy and health data protection, data de-identification and reidentification, drug detailing (marketing), commercial benefit from the required disclosure of personal information, clinician privacy and the duty of confidentiality, beneficial and unsavory uses of health data, regulating health technologies, and considering data as speech. Individuals should, at the very least, be aware of how data about them are collected and used. Taking account of how those data are used is needed so societal norms and law evolve ethically as new technologies affect health data privacy and protection.*

**Kaplan, Bonnie. 2016. "How Should Health Data Be Used? Privacy, Secondary Use, and Big Data Sales." *Cambridge Quarterly of Healthcare Ethics* 25 (2):312-329. doi: 10.1017/S096318011500061**

*Electronic health records, data sharing, big data, data mining, and secondary use*

*are enabling exciting opportunities for improving health and healthcare while also exacerbating privacy concerns. Two court cases about selling prescription data, the Sorrell case in the U.S. and the Source case in the U.K., raise questions of what constitutes "privacy" and "public interest"; they present an opportunity for ethical analysis of data privacy, commodifying data for sale and ownership, combining public and private data, data for research, and transparency and consent. These interwoven issues involve discussion of big data benefits and harms and touch on common dualities of the individual versus the aggregate or the public interest, research (or, more broadly, innovation) versus privacy, individual versus institutional power, identification versus identity and authentication, and virtual versus real individuals and contextualized information. Transparency, flexibility, and accountability are needed for assessing appropriate, judicious, and ethical data uses and users, as some are more compatible with societal norms and values than others.*

**McEwen, Jean E., Joy T. Boyer, and Kathie Y. Sun. 2013. "Evolving approaches to the ethical management of genomic data." *Trends in Genetics* 29 (6): 375-382.**

*The ethical landscape in the field of genomics is rapidly shifting. Plummeting sequencing costs, along with ongoing advances in bioinformatics, now make it possible to generate an enormous volume of genomic data about vast numbers of people. The informational richness, complexity, and frequently uncertain meaning of these data, coupled with evolving norms surrounding the sharing of data and samples and persistent privacy concerns, have generated a range of approaches to the ethical management of genomic information. As calls increase for the expanded use of broad or even open consent, and as controversy grows about how best to handle incidental genomic findings, these approaches, informed by normative analysis and empirical data, will continue to evolve alongside the science.*

**Mittelstadt, Brent, and Luciano Floridi. 2016. "The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts." *Science & Engineering Ethics* 22 (2):303-341. doi: 10.1007/s11948-015-9652-2.**

*This article systematically and comprehensively analyses academic literature concerning the ethical implications of Big Data, providing a watershed for future ethical investigations and regulations. Particular attention is paid to biomedical Big Data due to the inherent sensitivity of medical information. By means of a meta-analysis of the literature, a thematic narrative is provided to guide ethicists, data scientists, regulators and other stakeholders through what is already known or*



*hypothesised about the ethical risks of this emerging and innovative phenomenon.*

**Ossorio, Pilar N. "The Ethics of Translating High-Throughput Science into Clinical Practice." *Hastings Center Report* 44, no. 5 (2014): 8-9.**

*Biomedical research is increasingly data intensive and computational, and "big data science" is migrating into the clinical arena. Unfortunately, ethicists, regulators, and policy-makers have barely begun to explore the ethical, legal, and social issues raised by the variety of analytical and computational approaches in use and under development in biology and medicine. Most scholarship concerning big data bioscience has focused on privacy, a vitally important consideration but not the only one. Among the issues raised by new computational technologies are questions about safety and safety assessment, justice, and how to obtain proper informed consent. These technologies also raise a myriad of regulatory issues that could influence the probability of translating new assays or computational tools to the clinical or public health spheres.*

**Ozdemir, V., and E. Kolker. 2016. "Precision Nutrition 4.0: A Big Data and Ethics Foresight Analysis--Convergence of Agrigenomics, Nutrigenomics, Nutriproteomics, and Nutrimetabolomics." *OMICS* 20 (2):69-75. doi: 10.1089/omi.2015.0193.**

*The authors propose the creation of a new Precision Nutrition Evidence Barometer for periodic, independent, and ongoing retrieval, screening, and aggregation of the relevant life sciences data. For innovation in Big Data ethics oversight, the authors suggest "nested governance" wherein the processes of knowledge production are made transparent in the continuum from life sciences and social sciences to humanities, and where each innovation actor reports to another accountability and transparency layer: scientists to ethicists, and ethicists to scholars in the emerging field of ethics-of-ethics. Such nested innovation ecosystems offer safety against innovation blind spots, calibrate visible/invisible power differences in the cultures of science or ethics, and ultimately, reducing the risk of "paper values"--what people say--and "real values"--what innovation actors actually do. The authors are optimistic that the convergence of nutrigenomics with nutriproteomics, nutrimetabolomics, and agrigenomics can build a robust, sustainable, and trustworthy precision nutrition 4.0 agenda, as articulated in this Big Data and ethics foresight analysis.*

**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.*

**Poldrack, Russell A., and Krzysztof J. Gorgolewski. 2014. "Making big data open: data sharing in neuroimaging." *Nature Neuroscience* 17 (11):1510-1517. doi: 10.1038/nn.3818.**

*In the last decade, major advances have been made in the availability of shared neuroimaging data, such that there are more than 8,000 shared MRI (magnetic resonance imaging) data sets available online. Here we outline the state of data sharing for task-based functional MRI (fMRI) data, with a focus on various forms of data and their relative utility for subsequent analyses. We also discuss challenges to the future success of data sharing and highlight the ethical argument that data sharing may be necessary to maximize the contribution of human subjects.*

**Prainsack, Barbara. 2015. "[Three 'H's for Health -- The Darker Side of Big Data.](#)" *Bioethica Forum: Schweizer Zeitschrift für Biomedizinische Ethik (Swiss Journal of Biomedical Ethics)* 8 (2):40-41.**

*In September 2015, a woman made the news because while trying to take her child to the hospital, her bank remotely deactivated the ignition to her car because she was behind on her mortgage payments. The author uses this case to discuss how the concepts of hypercollection, harm and humiliation need to be kept in mind for designers of systems that use big data.*

**Purcell, Ryan H., and Karen S. Rommelfanger. 2015. "Internet-Based Brain Training Games, Citizen Scientists, and Big Data: Ethical Issues in Unprecedented Virtual Territories." *Neuron* 86 (2):356-359. doi: 10.1016/j.neuron.2015.03.044.**

*Internet brain training programs, where consumers serve as both subjects and*

*fundamentals of the research, represent the closest engagement many individuals have with neuroscience. Safeguards are needed to protect participants' privacy and the evolving scientific enterprise of big data.*

**Rothstein, Mark A. 2015. "Ethical Issues in Big Data Health Research: Currents in Contemporary Bioethics." *Journal of Law, Medicine & Ethics* 43 (2):425-429. doi: 10.1111/jlme.12258.**

*The article discusses the moral and bioethical aspects regarding big data health research in the U.S., including in regard to personal autonomy, medical informed consent and privacy. An overview of the 2009 report "Beyond the Privacy Rule: Enhancing Privacy, Improving Health through Research," published by the U.S. Institute of Medicine (IOM) is provided.*

**Tractenberg, Rochelle E. 2016. "Creating a Culture of Ethics in Biomedical Big Data: Adapting 'Guidelines for Professional Practice' to Promote Ethical Use and Research Practice." In *The Ethics of Biomedical Big Data*, edited by Daniel Brent Mittelstadt and Luciano Floridi, 367-393. Cham: Springer International Publishing.**

*This article looks at the concept of responsible research and innovation and responsible conduct of research in the field of "biomedical big data" and how the former is more likely to influence the ethical use of big data in this field. The article reviews guidelines from the American Statistical Association, the Royal Statistics Society, and the International Statistics Institute as well as the Association for Computing Machinery and shows how these can be leveraged to help researchers deal with the challenges currently arising from big data use in the biomedical field.*

**Vayena E, Marcel Salathé, Lawrence C. Madoff, and John S. Brownstein. 2015. [Ethical Challenges of Big Data in Public Health](#). *PLoS Comput Biol* 11(2): e1003904. doi:10.1371/journal.pcbi.1003904**

*In this article the authors identify some of the key ethical challenges associated with digital disease detection activities and outline a framework for addressing them. We argue that it is important to engage with these questions while the field is at an early stage of evolution in order to make ethical awareness integral to its development.*

**Watson, R. William G., Elaine W. Kay, and David Smith. 2010. "Integrating biobanks: addressing the practical and ethical issues to deliver a valuable tool for cancer research." *Nature Reviews Cancer* 10 (9): 646-651.**

*Cancer is caused by complex interactions between genes, environment and lifestyles. Biobanks of well-annotated human tissues are an important resource for studying the underlying mechanisms of cancer. Although such biobanks exist, their integration to form larger biobanks is now required to provide the diversity of samples that are needed to study the complexity and heterogeneity of cancer. clear guidelines and policies are also required to address the challenges of integrating individual institutional or national biobanks and build public trust. This Science and Society article highlights some of the main practical and ethical issues that are undergoing discussion in the integration of tissue biobanks for cancer.*

**Watson, R. William G., Elaine W. Kay, and David Smith. 2010. "Integrating biobanks: addressing the practical and ethical issues to deliver a valuable tool for cancer research." *Nature Reviews Cancer* 10 (9): 646-651.**

*Cancer is caused by complex interactions between genes, environment and lifestyles. Biobanks of well-annotated human tissues are an important resource for studying the underlying mechanisms of cancer. Although such biobanks exist, their integration to form larger biobanks is now required to provide the diversity of samples that are needed to study the complexity and heterogeneity of cancer. clear guidelines and policies are also required to address the challenges of integrating individual institutional or national biobanks and build public trust. This Science and Society article highlights some of the main practical and ethical issues that are undergoing discussion in the integration of tissue biobanks for cancer.*

**Zarate, Oscar A., Julia Green Brody, and Phil Brown. 2016. "Balancing Benefits and Risks of Immortal Data: Participants' Views of Open Consent in the Personal Genome Project." *Hastings Center Report* 46 (1):36-45.**

*In considering what informed consent and research protocols should entail in this era of big data, researchers and institutional review boards can benefit from listening to the experiences of people who volunteer for studies with extensive data sharing and novel forms of consent.*

## **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)**

Life and Environmental Sciences