Ethics in the Era of Big Data

9

Diego Librenza-Garcia

9.1 Initial Considerations

Technology and its consequences in human behavior and relationships have been fascinating mankind for centuries. A whole new literary genre was created with science fiction so that we could imaginatively explore what the future may hold for our species. Since then, both movies and novels have increasingly focused on technological advancement, == most often in dystopic scenarios, in which artificial intelligence creates prejudice and ethical dilemmas through biased handling of personal and collective data. Despite these catastrophic predictions, technological progress* has redefined our civilization and our way of life with exponential advances, to the point that some publications, such as The Economist, declared that data might be considered for this century what oil was to the last one, conceiving a whole new economic scenario (Economist 2017). In medicine, and more particularly in psychiatry, big data analytics represent a new era in which we are shifting from group-level evidence, as proposed by medicine-based evidence, to individual and personalized predictions, potentially leading to personalized care (Greenhalgh et al. 2014; Passos et al. 2016). Nevertheless, despite all prospects regarding the growth, sharing and processing of data, and all the benefits it may represent, this revolution does not come without risks.

D. Librenza-Garcia (✉)

Department of Psychiatry and Behavioural Neurosciences, McMaster University, Mood Disorders Program, Hamilton, ON, Canada

Graduation Program in Psychiatry and Department of Psychiatry, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil
e-mail: librenzagarcia@gmail.com

Although data, per se, is ethically neutral, what one decides to do with it may not be. Estimates point that in 2018, 50% of business ethics violations may happen through improper handling of these large data sets and its analysis (Herschel and Miori 2017). As most revolutions go, we are noticing both the benefits and the problems related to Big Data as it unfolds, and most of the time, by seeing its negative consequences and reacting to them rather than acting proactively.* However there is an optimistic* view of how big data and techniques such as machine learning may improve health services in all respects. (Barrett et al. 2013; Angus 2015; Insel and Cuthbert 2015; Huys et al. 2016; Beam and Kohane 2018). Not only can this* improve hospital and doctor performance, but also an individual's quality of life and how patients understand and interact with these disorders (or the perspective of presenting them in the future). On the other hand, we are unaware of how big data may negatively impact* these same dimensions or create new types of inequality.

The present chapter provides a perspective on the ethical issues that may emerge from big data analytics and how this may challenge us in the coming years. Although ethics may have many definitions that go than "what is right and what is wrong," an ongoing field must adapt to new realities as well as the ethical issues and the discussion of how to deal with them may have many definitions that go well beyond "what is right and what is wrong"* paramount (Davis 2012). In fact, we are already experiencing the impact of big data for many years now and may see its influence exponentially increasing in the next years. For this chapter, we chose to divide the ethical challenges into four sections. First, regarding the data itself and its handling. Second, the impact that predictive models created with this data may have for patients. Third, the ethical issues created by these same models to clinicians, and fourth, the ethical issues involved in research, especially regarding informed consent.

9.2   Ethical Issues Regarding Data

Data has been created since the beginning of civilization, first in the form of pictures drawn by our ancestors in caves, then by written registers and, nowadays, created, stored and processed by a myriad of electronic devices that are continually registering and creating information (World Economic Forum 2011; Lantz 2015; Beam and Kohane 2018). What changed recently is the speed at which we create and store data and the fact that now we have both the methods and the computational capacity to extract useful insights from this vast amount of information (Lantz 2015). However, from the collecting to the application of this massive flow of data, some questions arise. Who owns this information, and how can it be used? How may this constant flow harm individual privacy, or how may a lack of transparency facilitate a data monopoly, where a minority of individuals may consolidate power and control? of transparency generate. The legislation is still emerging and many of these questions remain open to discussion, and we are probably looking at two opposing risks. First, that data may be poorly handled and create negative

consequences for individual and society; and second, that the perception of this threat may lead to disproportional overregulation that could slow down and delay the positive effects of big data.

9.2.1 Privacy and Anonymity

It is hard to think of any human activity nowadays that does not generate data, given how connected we are with electronic devices and, in consequence, interconnected with each other. Our behavior produces a data imprint, which may allow others to detect our behavioral patterns, and reveal our personal preferences* (Davis 2012; Murdoch and Detsky 2013). Although terms of service from software that collect personal data usually mention privacy and assure data anonymity, they can sometimes be vague and superficial in their description. In some cases, one can analyze this "anonymized data" and through reverse engineering, trace information back to a singular individual, a process called re-identification (Tene and Polonetsky 2013; Mello et al. 2013; Terry 2014). This precedent is of extreme importance in the medical setting, as health-related data may contain sensitive information about the patients, such as sexual orientation, previous history of abortions, suicide attempts and so on. Moreover, patients are vulnerable because of their expectations regarding their diagnosis or apprehension towards treatment and prognosis, and the disclosure of this information may complicate even more how they experience their disorder or treatment.

It is also essential to determine who should have access to data and for what purpose. Re-identification or hackings may lead to data leakage and exposure of sensitive information, but physical and remote access to stored data may also give an individual opportunity of duplicating a data set and releasing this information (Culnan and Williams 2009). Those who are granted direct access* to the data and handle it in their daily work are in a position of power. Companies and institutions need to establish clear policies to determine who is granted access to this information, to avoid sensitive data to be inadequately visualized, analyzed and exposed (Davis 2012).

Given how dynamic Big Data is, it is almost impossible to actively monitor how private information is being stored and propagated. Agreement terms that indicate that data will be used to "personalize experience" or "improve performance" may fail to inform, for example, if that data is being sold or transferred to third parties—a widespread practice—and for what those third parties may use it. The same information may have very different uses: one can create models based on social media information for very different tasks, such as selling a product or predicting harmful behavior such as suicide attempt. When an individual agrees to share their data, how exactly and for what these data is used are questions that either remain unanswered or are answered without the pertinent specificity. In the particular case of social media, although information is in a public virtual space, people may be unaware of its multiple uses and the commercial value of what they are producing. Lastly, there is a risk that anonymized data may be clustered according

to geographical, ethnic or last sexual orientation, that may lead to discrimination and stigmatization—in this case, affecting not only the individuals that share their data but also others in these clusters (Craig and Ludloff 2011; Schadt 2012; Mittelstadt and Floridi 2016).

9.2.2 Ownership

Since we are unceasingly producing data, which is continuously being stored, who does this data belong to exactly? It is unthinkable that all this information can be managed by the individual that generates it across the unending stream of information that goes from our devices to corporations and governments, and then back to the individual in the form of actions or products. How much value can be assigned to a given amount of information, and can a corporation sell a given individual's personal data? It is somewhat disturbing that someone might own people's personal information, as well as their behavior and preferences, and may employ these to influence future behavior and preferences. The boundaries here are also uncertain: which data may be public and which data may remain private? Which data may lie in between, accessible for purposes of research and innovation but not entirely public? From the moment a patient enters an emergency room until its discharge several days later, he generates a variety of data. Should the institution be free to use all kinds of data, some of them, or none, and whom may have access to the raw data and the insights extracted from it? It is unarguable how useful this information is, but there are no universal regulations on the matter. Furthermore, ownership may be defined not by only possessing the rights to compile and use with exclusivity the data, but also to the right of analyzing and use it to create new technologies, generating copyrighted products or patents (Choudhury et al. 2014).

9.2.3 Transparency

Data gathering services should not only be transparent about what they are collecting and what are the potential uses of the data, but they should state this in a clear and concise way. A study found that, if one stops to read each term of agreement in a year, one would waste approximately 76 work days reading them (McDonald and Cranor 2008). When an individual is sharing his data, it is relevant to know the ethical principles of the institution in charge of the data gathering, what they intend to do with the information and what is out of boundaries (Davis 2012; Liyanage et al. 2014). In recent years, we have seen many cases in which data was secretly collected and analyzed, and with no purpose known to the users of the service (van der Sloot 2015). Despite the violation of the individual autonomy, this course of action may discourage people from sharing their data even in reliable and transparent platforms, thus limiting the data available for analysis. As already pointed out, it should be clear if the data set would be shared with third parties, or sold to them, or even aggregate external sources.

9.2.4 Identity and Reputation

Technological advancements have altered the way we see ourselves as individuals. Nowadays, our identity consists of both our offline and online activities, and our reputation is influenced by our behavior in both these dimensions. Our offline behavior may impact our online reputation and vice-versa (Davis 2012; Andrejevic 2014). In this sense, the possibility of sensitive data exposure as a result of re-identification or hacking may have an impact in the offline and online parts of people's identity, and therefore harming their reputation. It is not clear how some platforms deal with sensitive in some cases and how much it is protected. Even if agencies with highly classified information are hacked, it is worrisome to think how vulnerable other information may be, such as electronic records or private files. A breach of privacy, therefore, may lead to irreversible and harmful repercussions in how we and others perceive ourselves.

9.2.5 Reliability

Beyond the traditional "3 V's" of Big Data—Variety, Velocity, Volume—IBM proposed a fourth V, veracity (Zikopoulos et al. 2012). Data is not always reliable—it could be human error or bias when a person is collecting the data, or perhaps the use of an uncalibrated device that gives wrong measures, or just the fact that subjects of interest may opt-out, with loss of relevant information. The analysis of incomplete, biased or out of context data may lead to incorrect conclusions, and those conclusions may lead to harmful action or decisions (Bail 2014; Markowetz et al. 2014). Moreover, data is increasingly becoming collected autonomously, by sensor devices, and not infrequently, being processed and analyzed independently of human interference also. The complexity of algorithms used in this analysis—the so called black box methods—may result in our inability to understand how they work, which is troublesome when these same algorithms may be used to influence behavior or make decisions with high impact on one's treatment and prognosis, for example (Lantz 2015).

We should avoid models that are biased in nature. For example, when creating an algorithm to predict suicide attempts, via collecting social media data, users may not be representative of those who use another platform, or those who, although have an account, are not active. Although it may be argued that not being active is also valuable information, this model will fail to identify suicide attempts among inactive individuals of this network, that may be generating relevant data in another platform that is relevant to the topic of interest. On the other hand, a universal model including all internet-related information plus offline use of devices for the individual may be closer to the aim of predicting suicide—although with higher costs and astounding complexity. Before applying any algorithm in real life

scenarios, we should take these problems into account, to prevent that biased models with incorrect or incomplete conclusions ended up causing more harm than benefit (Andrejevic 2014).

9.3   Ethical Issues Regarding Patients

Predictive psychiatry may contribute to improve outcomes and prevent disability or harm, but it may also produce harm, influencing other spheres beyond individual's health. If we can predict that an individual will have a more pernicious illness course, that would mean he will make more use of health services, and therefore, may be charged more for a health plan. The prediction, per se, may not be an issue, but the application may perhaps be. For instance, it may be possible that unfavourable outcomes of an individual may fuel eugenic policies or even create social prejudice regarding the subjects with these outcomes.

We should also worry about how devastating a prediction could be. One classic example is Huntington's disease, an autosomal dominant disorder that can be predicted by a simple genetic test. A positive test may tell a patient that he will, in the next years, experience a progressive and severe loss of its brain functions, while the subject is still healthy. If an individual is predicted to develop a psychiatric disorder years before its onset, how many this information influence his quality of life, or ability to avoid that outcome? How will it influence his relationships with his peers or change the course of his actions in the scenario where he was not informed of the outcome? It is possible that the stressful burden of knowing may incur in speeding the disorder installment or even lead to another disorder, such as a depressive episode or substance abuse, in the prior years before the onset of the predicted disorder. A question of the uttermost importance in big data ethics is how our patients may cope with such predictions about their future, and weigh harm and benefit of its use. It is different if we develop an intervention to prevent the outcome and can offer it to an individual. The following clinical cases illustrate some of these ethical dilemmas.

Case 1
J. is an 18-year old male who decides to enlist and serve in the Army. After collecting a series of clinical data and undergo neuroimage acquisition and analysis of serum biomarkers, he is predicted to develop PTSD along with a mood disorder during his time serving with 98% accuracy. Moreover, the algorithm also predicted with an accuracy of 92% that he would attempt suicide in the following year. He still wants to serve the Army even knowing the risks. However, he is then dismissed against his will.

Case 2

C. is a 15-year old female whose father has bipolar disorder with a pernicious trajectory marked by functional impairment and disability, as well as metabolic disorders. At the will of her mother, she underwent a test that can predict with almost 100% accuracy if one will develop a psychiatric disorder in the future. She is then predicted to develop bipolar disorder with a similar course of her father in the next ten years. There is no available treatment at the time to prevent this conversion.

Although big data analytics may have several benefits and a substantial social impact to prevent outcomes such as PTSD, one may argue that there is no absolute prediction and that the individual may have the autonomy to choose to serve the army regardless. However, from a legal perspective, enlist an individual with high chances of developing a debilitating disorder may incur in health-care related expenses and pensions. Moreover, if he develops a disorder on the battlefield, it is possible that his symptoms may jeopardize his safety and that of other soldiers. There is also a possibility of joining the Army but not be sent to the field—which may stigmatize J. as being unable for some medical reason to go to combat.

In the second scenario, knowing that C. will most likely develop BD may help in screening her for the first symptoms of the disorder and allow early intervention when needed. She may start attending an outpatient clinic before the installment of the disorder. She will probably need familiar and professionalized support throughout this prodromal period. Again, there is a chance she will not develop the disorder cause the prediction is not perfectly accurate, and she may undergo all this traumatic experience unnecessarily. Also, as she is a minor, should her mother decide she does not need to know at this point, what course of action should the psychiatrist take?

What is common to both cases is the uncertainty of the prediction. It is hard to imagine a 100% accurate application to predict an outcome, at least with our current state-of-the-art resources. There is always the possibility of that outcome not happening, and the individual forced to live with the burden of its possibility. Although most algorithms and models in current studies are still in proof-of-concept phases so far, it is possible that patients should experience this dilemma in the future. In this uncharted territory, there is no delimited policy or guidelines on how to proceed, nor protocols available for follow-up and assessment. Medical guidelines may have to address the problem of "potential patients," that do not manifest any symptoms at the time of the prediction.

9.4   Ethical Issues Regarding Clinician Decision

We can hypothesize at some point in the future, machines may provide diagnoses with better accuracy than physicians, as some algorithms are already achieving higher accuracies with machine learning than doctors to diagnose certain conditions

(Liu et al. 2017). They can also be used to redefine diagnosis by grouping patients with similar characteristics and integrating different levels of information in such a convoluted way that the meaning of this categories may be impossible for us to understand (Insel and Cuthbert 2015; Huys et al. 2016). The positive implications include predicting treatment response or detecting a disorder* before its onset and may alert us which patients will experience unfavorable functional or cognitive outcomes and have a more severe illness course (Passos et al. 2016; Librenza-Garcia et al. 2017). Predictive models open a door not only to prevention of these outcomes due to early intervention strategies but also to efforts to avoid conversion to a disorder. Amidst all these advances, the clinician finds himself as a bridge between patient and machine, trying to deal with patient expectations and technological insights.

Technology, however, is still dependent on our input. We have to define a psychiatric disorder and the outcome for the machine to interpret, and if we do it wrong, all data and inferences about it would be, in consequence, useless. Machines could get insight on data that we cannot, but we still need to interpret its findings. We can data mine for clusters of patients and redefine the way we diagnose, but given the number of different ways this could go, we should still choose which road we will take from there. At least in psychiatry, it is unimaginable—for now—to think that a machine could replace the clinician, given the importance of empathy and the doctor-patient relationship. The two cases below illustrate some challenges in clinician decision.

Case 3

A psychiatrist will discharge an inpatient after a month of hospitalization. He performs a standard battery of exams and gather clinical data and uses a phone application that can predict suicide attempt in the next three months with high accuracy. Despite being euthymic and with no suicidal ideation at the time, the patient is predicted to attempt suicide in this period.

Case 4

After a series of appointments in an outpatient clinic, the psychiatrist evaluating F. gives him a diagnosis of major depressive disorder. By gathering genetic, neuroimaging, clinical and serum biomarkers data, an algorithm predicts with a high accuracy that the patient has, in fact, bipolar disorder. The psychiatrist, then, reconsiders his choice of monotherapy with an antidepressant.

It is very likely that predictions may impact on clinician decision. If the patient in case 3 is predicted to attempt suicide, should he stay in inpatient care for a greater amount of time, or go home with familiar surveillance and regular appointments? If he lives alone, should he receive domiciliary follow-up as well? If by one side this prediction may provide better resource assignments for those predicted to attempt suicide, it can also lead to neglection of those predicted not to undergo this outcome. Since no model is perfect, some of the high-risk individuals may receive a regular follow-up, and the clinician may relax and neglect important risk signs, reassured by the negative prediction. In the case of F., despite the clinical diagnosis, the psychiatry may be reluctant cause the depressive episode may be only a first manifestation of bipolar disorder and may be followed by a manic presentation in the future—in the worst-case scenario, an iatrogenic manic switch triggered by his choice of treatment. On the other hand, if the prediction is wrong, he may be depriving the patient of a first line treatment and using an additional and unnecessary mood stabilizer, with all its known side effects.

9.5   Ethical Issues in Research

Informed consents in psychiatric research are usually developed stating what data will be collected and to what end. This poses a challenge because one of the purposes of big data analytics is to extract new knowledge or patterns from that information, ones that may not be included in the initial aim of a study—especially if we are dealing with unsupervised models. So, it is a challenge on how to include the unpredictable in the informed consent. Patients usually consent to participate in a single study, but big data may be more useful if data is shared, integrated and reanalyzed between different groups, increasing its complexity but also providing us with even more useful insights (Ioannidis 2013; Larson 2013; Choudhury et al. 2014). Also, we usually do not state for patients if whatever insight we obtained from the data may result in any feedback to them. If we create a model to predict response to antidepressants that have high accuracy and applicability, and it predicts that a patient in the validate sample will relapse with the medication he is currently using, will he be informed? Although this sound logical, should we also inform a patient if the accuracy is relevant, but not applicable?

Another relevant question is how we should handle social media information. Although it may have been made public, is the individual aware that his information can be used in a health-related scenario? How should we gather consent in such a vast universe? (Krotoski 2012; Lomborg and Bechmann 2014). One may hypothesize that in the future an individual may "opt-in" to data in which he is willing to share, and for which application*, but for now, each platform, software or website has a different policy (Prainsack and Buyx 2013). Broader consent policy may resolve the issue on the end of big data but not of the individual while listing possible future uses and authorization for each may be more comfortable for the

patient but limit newer insights into that data in the future. Reassessment for new consent can also be one strategy, but it will probably reduce the sample due to follow-up losses (Currie 2013; Lomborg and Bechmann 2014). Moreover, it would increase the costs and bureaucracy and slow down or preclude future research.

The fact is, for most of our studies, informed consent was designed to tackle themes relevant to evidence-based medicine, with predefined questions and a limited amount of answers expected. From now on, it is necessary to find a way to adapt it to this new reality, which includes the uncertainty of what the data can reveal and how it can impact patients afterward.

9.6   Conclusion

In the past, we would not dare to dream how big data would defy our limits and see far beyond what we can, nor how it could expand the limits of the world by not only redefining the real world but also creating uncountable virtual ones. It is undeniable that Big data is pushing us to consider ethical issues and whether they violate fundamental, civil, social, political or legal rights. On the other hand, big data analytics will also redefine what we think is possible in the next few years, with the possibility of devices being even more ingrained in our daily patterns of behavior, through digital profiling, and artificially intelligent-driven politics. The aforementioned ethical issues are only the ones we are facing now and in the near future. New issues may arise in areas that do not even exist at this time, and more challenges will surface as big data technology continues to evolve and expand its influence in our lives. There is no telling how much we will advance and how far the possibilities of this evolution may lead us, and what unforeseen ethical issues may arise ahead. Whether big data and artificial intelligence will guide us towards a dystopic or utopic society, it depends on how we will handle these ethical issues from now on. Technology, like every resource, is primarily neutral and can be used to cause both benefit and harm.

There is a delicate balance that we shall seek for the sake of an efficient and human health-care. A lack of policies on how to handle and utilize data may result in more inequality and create unpredictable harm to society and individuals. Nevertheless, if society lets itself to be driven by unfounded concerns about these new technologies, it may overreact and create preemptive obstacles, to the point in which a restrictive and overregulated policy may prevent not only harm but also progress and benefits that could improve patient care and change illness' course.

Some of the values we have today may evolve as new challenges arrive, which will promote a reformulation of our ethical principles. In this fashion, big data ethics do not consist of absolute and immutable principles, but, on the opposite, it is malleable according to the challenges and outcomes not prior anticipated. Some

scenarios presented in this chapter are already challenging, and there is no telling what new ones may lie ahead. Nevertheless, besides all potential innovations and problematic scenarios big data may cause, one fundamental principle of medicine stated in the Hippocratic Oath still applies: *primum non nocere (First, to do no harm)*.

References

Andrejevic M (2014) The big data divide. Int J Commun 8:1673–1689. 1932–8036/20140005

Angus DC (2015) (NIG) fusing randomized trials with big data: the key to self-learning health care systems? JAMA J Am Med Assoc 314:767–768. https://doi.org/10.1001/jama.2015.7762

Bail CA (2014) The cultural environment: measuring culture with big data. Theory Soc 43:465–524. https://doi.org/10.1007/s11186-014-9216-5

Barrett MA, Humblet O, Hiatt RA, Adler NE (2013) Big data and disease prevention: *from quantified self to quantified communities*. Big Data 1:168–175. https://doi.org/10.1089/big.2013.0027

Beam AL, Kohane IS (2018) Big data and machine learning in health care. JAMA J Am Med Assoc 319:1317–1318. https://doi.org/10.1001/jama.2017.18391

Choudhury S, Fishman JR, McGowan ML, Juengst ET (2014) Big data, open science and the brain: lessons learned from genomics. Front Hum Neurosci 8:1–10. https://doi.org/10.3389/fnhum.2014.00239

Craig T, Ludloff ME (2011) Privacy and big data: the players, regulators, and stakeholders. O'Reilly Media

Culnan MJ, Williams CC (2009) How ethics can enhance organizational privacy: lessons from the choicepoint and TJX data breaches. MIS Q 33:673–687. https://doi.org/10.2307/20650322

Currie J (2013) "Big data" versus "big brother": on the appropriate use of large-scale data collections in pediatrics. Pediatrics 131:S127–S132. https://doi.org/10.1542/peds.2013-0252c

Davis K (2012) Ethics of big data: balancing risk and innovation. O'Reilly Media

Economist T (2017) The world's most valuable resource is no longer oil, but data. Econ

Greenhalgh T, Howick J, Maskrey N (2014) Evidence based medicine: a movement in crisis. BMJ 348:g3725–g3725. https://doi.org/10.1136/bmj.g3725

Herschel R, Miori VM (2017) Ethics & big data. Technol Soc 49:31–36. https://doi.org/10.1016/j.techsoc.2017.03.003

Huys QJM, Maia TV, Frank MJ (2016) Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci 19:404–413. https://doi.org/10.1038/nn.4238

Insel TR, Cuthbert BN (2015) Brain disorders? Precisely. Science 348:499–500. https://doi.org/10.1126/science.aab2358

Ioannidis JPA (2013) Informed consent, big data, and the oxymoron of research that is not research. Am J Bioeth 13:40–42. https://doi.org/10.1080/15265161.2013.768864

Krotoski AK (2012) Data-driven research: open data opportunities for growing knowledge, and ethical issues that arise. Insights UKSG J 25:28–32. https://doi.org/10.1629/2048-7754.25.1.28

Lantz B (2015) Machine learning with R - second edition. Cambridge University Press, Cambridge

Larson EB (2013) Building trust in the power of "big data" research to serve the public good. JAMA 309:2443. https://doi.org/10.1001/jama.2013.5914

Librenza-Garcia D, Kotzian BJ, Yang J et al (2017) The impact of machine learning techniques in the study of bipolar disorder: a systematic review. Neurosci Biobehav Rev 80:538–554. https://doi.org/10.1016/j.neubiorev.2017.07.004

Liu Y, Gadepalli K, Norouzi M, et al (2017) Detecting cancer metastases on gigapixel pathology images. 1–13. https://doi.org/10.1016/j.ejim.2017.06.017

Liyanage H, De Lusignan S, Liaw S et al (2014) Big data usage patterns in the health care domain: a use case driven approach applied to the assessment of vaccination benefits and risks contribution of the IMIA primary healthcare working group big data for assessing vaccination benefits and risks: A. IMIA. Yearb Med Inform:27–35

Lomborg S, Bechmann A (2014) Using APIs for data collection on social media. Inf Soc 30:256–265. https://doi.org/10.1080/01972243.2014.915276

Markowetz A, Błaszkiewicz K, Montag C et al (2014) Psycho-informatics: big data shaping modern psychometrics. Med Hypotheses 82:405–411. https://doi.org/10.1016/j.mehy.2013.11.030

McDonald AM, Cranor LF (2008) The cost of reading privacy policies. A J Law Policy Inf Soc 4:543–568

Mello MM, Francer JK, Wilenzick M et al (2013) Preparing for responsible sharing of clinical trial data. N Engl J Med 369:1651–1658. https://doi.org/10.1056/NEJMhle1309073

Mittelstadt BD, Floridi L (2016) The ethics of big data: current and foreseeable issues in biomedical contexts. Sci Eng Ethics 22:303–341. https://doi.org/10.1007/s11948-015-9652-2

Murdoch TBTB, Detsky ASAS (2013) The inevitable application of big data to health care. JAMA 309:1351–1352. https://doi.org/10.1001/jama.2013.393

Passos IC, Mwangi B, Kapczinski F (2016) Big data analytics and machine learning: 2015 and beyond. Lancet Psychiatry 3:13–15. https://doi.org/10.1016/S2215-0366(15)00549-0

Prainsack B, Buyx A (2013) A solidarity-based approach to the governance of research biobanks. Med Law Rev 21:71–91. https://doi.org/10.1093/medlaw/fws040

Schadt EE (2012) The changing privacy landscape in the era of big data. Mol Syst Biol 8:1–3. https://doi.org/10.1038/msb.2012.47

Tene O, Polonetsky J (2013) Big data for all: privacy and user control in the age of analytics

Terry N (2014) Health privacy is difficult but not impossible in a post-HIPAA data-driven world. Chest 146:835–840. https://doi.org/10.1378/chest.13-2909

van der Sloot B (2015) How to assess privacy violations in the age of big data? Analysing the three different tests developed by the ECtHR and adding for a fourth one. Inf Commun Technol Law 24:74–103. https://doi.org/10.1080/13600834.2015.1009714

World Economic Forum (2011) Personal data: the emergence of a new asset class

Zikopoulos PC, DeRoos D, Parasuraman K, et al (2012) Harness the power of big data