



The Ethics of Biomedical ‘Big Data’ Analytics

Brent Mittelstadt¹

Published online: 8 March 2019
© Springer Nature B.V. 2019

1 Introduction: The Ethics of Biomedical Data Analytics

In modern information societies, individuals generate streams of diverse and potentially valuable data. Digital technologies now easily and routinely record data about the behaviours and preferences of individuals at an unprecedented scale. Analytic techniques to make sense of this glut of data have grown in parallel, ushering in what some have called the age of ‘Big Data’.

Data analytics at scale provide huge opportunities to improve private and public life, especially in the health sector. In biomedical research and development, the analysis of large datasets (or ‘Biomedical Big Data’; henceforth BBD) has become a major driver of innovation and success, with partnerships between private data-intensive firms and public health bodies increasingly common (Powles and Hodson 2017). Potentially insightful health-related data can now be generated via social media applications and health platforms (Lupton 2014; Costa 2014), emerging ‘personal health monitoring’ technologies (Mittelstadt et al. 2014), home sensors (Niemeijer et al. 2010) and smart phone applications, and online forums and search queries. These new data sources complement traditional repositories consisting of aggregated clinical trials (Costa 2014), genetic and microbiomic sequencing data (McGuire et al. 2008), biological specimens, electronic health records, and administrative hospital data.

The value of Big Data analytics stems from the seemingly unlimited opportunities now available to link, aggregate, and search across diverse datasets to identify ‘small patterns’ or connections between phenomena and people (Floridi 2012). BBD provides new ways of understanding health and well-being at the level of the individual and society, for example by predicting behaviours, monitoring diseases and outbreaks, and providing risk stratification for individual patients. Possible applications include development of clinically useful predictive models (Choudhury et al. 2014), longitudinal and cross-sectional assessment of the effectiveness and efficiency of interventions and organisations (Tene and Polonetsky 2013), and longitudinal monitoring of chronic

✉ Brent Mittelstadt
brent.mittelstadt@oii.ox.ac.uk

¹ Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK

conditions and well-being (Boye 2012). Epidemiology (Mittelstadt et al. 2017), infectious disease research, and genomics and genetics (Heitmueller et al. 2014; Kaye et al. 2012) are already deeply affected.

While huge potential exists to advance the diagnosis, treatment, and prevention of diseases as well as foster healthy habits and practices (Costa 2014), the inherent sensitivity of health-related data and the implicit vulnerability of patients (Pellegrino and Thomasma 1993) pose ethical risks which cannot be ignored. The unprecedented volume and variety of data now available to these sectors challenge accepted social, ethical, and professional norms. Further, the growing reliance on algorithms to analyse them and to reach decisions, and the gradual reduction of human oversight over many automatic processes pose pressing issues of fairness, responsibility, and respect of human rights. As is often the case with rapid scientific and technological progress, understanding of these challenges lags behind.

These issues can be addressed successfully. However, if they are overlooked, underestimated or left unresolved, they risk hindering the innovation and progress that BBD can bring to society at large and to future generations. Furthermore, as recent events involving the NHS care.data programme show, BBD projects may face a double bottleneck: ethical mistakes or misunderstandings may lead to social rejection or distorted legislation and policies, which in turn may cripple the acceptance and advancement of data science. Similar to the public debate over genetically modified organisms (Devos et al. 2008), potentially beneficial projects may be put at risk through association with problematic applications.

Ethical foresight, incorporated at all stages of BBD initiatives, can help distinguish the good from the bad. Attention must be paid to known issues with Big Data analytics, covering topics such as informed consent, privacy, confidentiality, diversity, data ownership, digital divides, collective rights, and inclusive governance of research data (Mittelstadt and Floridi 2016; Mittelstadt 2017; Taylor, Floridi, and van der Sloot 2017). Proactive research and governance addressing these issues can help to understand impact, anticipate risks and unethical consequences, suggest early interventions to avoid or mitigate them, foster resilience, reinforce ethical goals and outcomes, and ensure that ethical best practices are developed, implemented, and appreciated.

To contribute to this critical step, this special issue of *Philosophy and Technology* aims to map new, under-researched but important issues, concepts, and cases that should inform proactive ethical assessment of emerging BBD applications and services. The papers contained within map and critically assess the current and potential ethical challenges facing Big Data in biomedicine.

To begin, the impact of personal health-monitoring devices on user autonomy and agency is examined by two contributions. John Owens and Alan Cribb argue that personal health devices such as the ‘FitBit’, which claim to help users live healthier lives by monitoring behaviour and feeding back information to promote healthy decisions, may instead expose users to risks of anxiety, stigma, and reinforcement of health inequalities. Despite providing potentially useful information, the authors are dubious of the devices’ actual contribution to user autonomy in terms of controlling or improving their health. To do so, they distinguish between procedural (or deliberative) and relational (or action-oriented) notions of autonomy. On the surface, wearable technologies seem to only provide information that could potentially be useful for making decisions

about one's lifestyle, while doing little to enhance actual opportunities to act to improve health.

In contrast, Nils-Frederic Wagner introduces the notion of 'patency' as a correlate to user agency. Health-monitoring devices are often thought to persuade or nudge users paternalistically towards health-promoting behaviours, which would seem to undermine the user's agency and autonomy. However, by employing the lens of the extended mind and extended will framework, Wagner argues that this portrayal of mHealth is misleading. While mHealth may render the agent passive through the receipt of technological commands, patency should be viewed not as a foil to agency, but rather as a correlate. From this perspective mHealth can simultaneously promote patency by nudging behaviour, while also serving as an effective technological tool to enhance user agency.

Concerns about autonomy are also reflected in four contributions looking at the impact of BBD on consent, trust, and data governance across different application areas. J. Patrick Woolley examines the role of trust and justice at a general level in biomedical Big Data analytics. Trust is often cited as a key value in data governance policy and oversight mechanisms, yet is often poorly grounded in a philosophical sense. Woolley argues that this is a key gap in existing scholarship, as different approaches to trust align differently with policy and governance structures. He unpacks how different philosophical notions of trust relate to traditional bioethical concepts and related laws, and their impact on striking a fair balance between individual and group interests in the sharing and re-use of data in BBD.

Chiara Garattini and her co-authors similarly seek to discuss ethical governance of BBD generally through lessons learned from a particular application area: infectious disease. They argue that BBD in infectious disease research and management is marked by new models of data accumulation which introduce four areas of ethical concern: the impact of (1) automation on autonomy, (2) complexity in Big Data analytics on informed consent, (3) profiling on identity and justice, and (4) greater population-level surveillance and interventions on behavioural norms and practices. Given the importance of these types of impact, proactive ethical assessment is urgently needed in infectious disease research and management to ensure responsible development, deployment, and societal acceptance of BBD.

Elvira Perez Vallejos and her co-authors reflect on their experiences with accessing online data from a youth web-counselling service for research. Digital mental health services pose particular ethical challenges for BBD due to the inherent sensitivity of the data in question. When coupled with a vulnerable user population, these challenges become particularly acute. The authors argue that particular attention must be given to the users' expectations of how their data will be re-used, specifically with regard to the perception of data as public, private, or open. Concrete recommendations for conducting online research involving vulnerable populations are proposed, including the need for a collaborative approach to data governance and access, and explicit opt-in and opt-out recruitment strategies.

Sebastian Schleidgen and his co-authors look at privacy risks in genomics research. Among the myriad data types utilised in BBD, genomics data poses unique risks of re-identification of participants. To help improve informed consent processes, the authors undertook a qualitative focus group in which patients and physicians at the National Center for Tumor Diseases were asked about their assessment of the informational risks

of participation in genomics research. The authors concluded that truly informed consent in genomics research requires (1) comprehensive disclosure of informational risks to participants, (2) independent governance entities, and (3) data sharing policies that offer guidance for physicians and researchers.

The final two contributions to the special issue examine the impact of new sources of information on equality and privacy in the delivery of medical care. Kristin Voigt considers how ‘social determinants of health’, which can generate inequalities in health outcomes, should be taken up by primary care providers. Information regarding the health of populations in specific geographical areas can increasingly be built from the ground up in BBD. Applying this population-level information in the care of individual patients can provide greater insight into social determinants of health, but may also pose privacy and equality risks when applied to individuals. Voigt provides a nuanced critique of the relationship between individual care and population-level medical knowledge to ensure BBD is deployed equitably in primary care.

Finally, Michele Loi examines the emergence of the ‘digital phenotype’, an extended human phenotype consisting of digital data that allows for medical conditions to be inferred and predicted (e.g. Tweets, Facebook posts, web search queries). This phenomenon allows for generalizable knowledge to be created from individual records and applied to others perceived to be similar. Loi suggests that ethical obligations are owed to individuals affected by this knowledge, and are not limited to individuals involved or identified in its creation. This philosophical critique of privacy and identity in the age of biomedical Big Data analytics thus stands in contrast to current privacy and data governance policy centred on the notion of ‘personal data’ linked to an identified or identifiable individual.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Boye, N. (2012). Co-production of health enabled by next generation personal health systems. *Studies in Health Technology and Informatics*, 177, 52–58.
- Choudhury, S., Fishman, J. R., McGowan, M. L., & Juengst, E. T. (2014). Big data, open science and the brain: lessons learned from genomics. *Frontiers in Human Neuroscience*, 8(239). <https://doi.org/10.3389/fnhum.2014.00239>.
- Costa, F. F. (2014). Big data in biomedicine. *Drug Discovery Today*, 19(4), 433–440. <https://doi.org/10.1016/j.drudis.2013.10.012>.
- Devos, Y., Maesele, P., Reheul, D., Van Speybroeck, L., & De Waele, D. (2008). Ethics in the societal debate on genetically modified organisms: a (re)quest for sense and sensibility. *Journal of Agricultural and Environmental Ethics*, 21(1), 29–61. <https://doi.org/10.1007/s10806-007-9057-6>.
- Floridi, L. (2012). Big data and their epistemological challenge. *Philosophy & Technology*, 25(4), 435–437. <https://doi.org/10.1007/s13347-012-0093-4>.
- Heitmueller, A., Henderson, S., Warburton, W., Elmagarmid, A., Pentland, A., & Darzi, A. (2014). Developing public policy to advance the use of big data in health care. *Health Affairs*, 33(9), 1523–1530. <https://doi.org/10.1377/hlthaff.2014.0771>.
- Kaye, J., Curren, L., Anderson, N., Edwards, K., Fullerton, S. M., Kanellopoulou, N., Lund, D., et al. (2012). From patients to partners: participant-centric initiatives in biomedical research. *Nature Reviews Genetics*, 13(5), 371–376. <https://doi.org/10.1038/nrg3218>.
- Lupton, D. (2014). The commodification of patient opinion: the digital patient experience economy in the age of big data. *Sociology of Health & Illness*, 36(6), 856–869. <https://doi.org/10.1111/1467-9566.12109>.

- McGuire, A. L., Colgrove, J., Whitney, S. N., Diaz, C. M., Bustillos, D., & Versalovic, J. (2008). Ethical, legal, and social considerations in conducting the human microbiome project. *Genome Research*, 18(12), 1861–1864. <https://doi.org/10.1101/gr.081653.108>.
- Mittelstadt, B. (2017). From individual to group privacy in big data analytics. *Philosophy & Technology*, 30(4), 475–494. <https://doi.org/10.1007/s13347-017-0253-7>.
- Mittelstadt, B., & Floridi, L. (2016). The ethics of big data: current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics*, 22(2), 303–341. <https://doi.org/10.1007/s11948-015-9652-2>.
- Mittelstadt, B., Ben Fairweather, N., Shaw, M., & McBride, N. (2014). The ethical implications of personal health monitoring. *International Journal of Technoethics*, 5(2), 37–60.
- Mittelstadt, B., Benzler, J., Engelmann, L., Prainsack, B., & Vayena, E. (2017). Is there a duty to participate in digital epidemiology? *Life Sciences, Society and Policy*.
- Niemeijer, A. R., Frederiks, B. J., Riphagen, I. I., Legemaate, J., Eefsting, J. A., & Hertogh, C. M. (2010). Ethical and practical concerns of surveillance technologies in residential care for people with dementia or intellectual disabilities: an overview of the literature. *International Psychogeriatrics*, 22, 1129–1142.
- Pellegrino, E. D., & Thomasma, D. C. (1993). *The virtues in medical practice*. New York: Oxford University Press.
- Powles, Julia, and Hal Hodson. 2017. 'Google DeepMind and healthcare in an age of algorithms'. *Health and Technology*, **March**, 1–17. <https://doi.org/10.1007/s12553-017-0179-1>.
- Taylor, Linnet, Luciano Floridi, and Bart van der Sloot, eds. 2017. *Group privacy: new challenges of data technologies*. 1st ed. Philosophical Studies. New York: Springer.
- Tene, Omer, and Jules Polonetsky. 2013. 'Big data for all: privacy and user control in the age of analytics'. *Nw.J. Tech. & Intell. Prop.* http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/nwteintp11§ion=20.