

# Introduction to the Special Theme: The expansion of the health data ecosystem – Rethinking data ethics and governance

Big Data &amp; Society

July–December 2019: 1–5

© The Author(s) 2019

Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/2053951719852969

[journals.sagepub.com/home/bds](https://journals.sagepub.com/home/bds)Tamar Sharon<sup>1</sup> and Federica Lucivero<sup>2</sup>

As in other domains, digital data are taking on an ever more central role in health and medicine today. And as it has in other domains, ‘datafication’ is contributing to a re-configuration of health and medicine, prompting its expansion to include new spaces, new practices, new techniques and new actors. Indeed, possibilities to quantify and ‘datafy’ areas of life that have not traditionally been considered the remit of biomedicine – such as sleep, ageing and emotions – and activities that have not traditionally been considered markers of health and disease – such as a person’s consumption patterns, her social media activity or her dietary habits – coupled with the promise of linking these heterogeneous datasets to glean medical insights, have contributed to a redefinition of almost any data as health-related data (Lucivero and Prainsack, 2015; Weber et al., 2014). Increasingly, these new types of data are being generated *outside* the traditional spaces of medicine, as people go about their daily lives interacting with consumer mobile devices. Similarly, the technological tools needed to capture, store, analyze and manage the flow of these data, from wearables and smart phones to cloud platforms and machine learning, increasingly rely on infrastructure and know-how that lie *beyond* the scope of traditional medical systems and scientists, amongst data scientists and information and communication technologies specialists. Moreover, new stakeholders are cropping up in these quasi-medical yet still undomesticated territories. On one end of the spectrum, individuals who generate health data as they track and monitor medical conditions, well-being, physical activity, or air quality, are both solicited as research participants and are making demands on researchers to utilize their personal health data (Health Data Exploration Project, 2014). On the other end of the spectrum, consumer technology corporations such as Apple and Google are reinventing themselves as obligatory passage points for data-

intensive precision medicine (Sharon, 2016). And somewhere in between, not-for-profit organizations, such as Sage Bionetworks and OpenHumans.org, are positioning themselves as mediators in this ecosystem in formation, *between* the medical research community, individual and collective generators of data and technology developers.

As proponents uphold, this expansion and decentralization of the health data ecosystem is promising, both in terms of the potential to advance data-driven research and healthcare, and in terms of rendering research more inclusive and more meaningful for participants (Shen, 2015; Topol, 2015). But, as critical scholars of science and technology have consistently shown, a fuller grasp of our technological present must always include the far-reaching, unexpected and sometimes deleterious social, political and cultural effects of discourses of scientific progress and technologically enabled democratization and participation. In recent years, such critical scholarship has been particularly wary of the new power asymmetries that datafication contributes to. Rather than levelling power relations, critics observe, these are being redrawn along new digital divides based on data ownership or access, control over digital infrastructures and new types of computational expertise, where those who generate data, especially citizens, patients and consumers, are positioned on the losing side of the on-going extraction and scramble for the world’s data driven by state

<sup>1</sup>Interdisciplinary Hub for Security, Privacy and Data Governance, Radboud University, Nijmegen, Netherlands

<sup>2</sup>Ethox and Wellcome Centre for Ethics and Humanities, University of Oxford, Oxford, UK

## Corresponding author:

Tamar Sharon, Radboud University, Nijmegen, Netherlands.

Email: [t.sharon@ftr.ru.nl](mailto:t.sharon@ftr.ru.nl)

This article is a part of special theme on Health Data Ecosystem. To see a full list of all articles in this special theme, please click here: [https://journals.sagepub.com/page/bds/collections/health\\_data\\_ecosystem](https://journals.sagepub.com/page/bds/collections/health_data_ecosystem).

and corporate actors (Andrejevic, 2014; boyd and Crawford, 2012; Taylor, 2017; Zuboff, 2015).

In the context of the data economy, the dominant response to these growing power differentials in regulatory, activist and technology development circles has been to ensure that individual data subjects acquire more control over the data they produce and how they flow – what Prainsack calls the ‘Individual Control’ approach in her contribution to this special theme. Examples include the EU’s General Data Protection Regulation, which confers data subjects more rights to control their personal data (e.g., data portability, the right to erasure), proposals to introduce property rights in personal data (Purtova, 2015), or initiatives that allow individuals to monetize their personal data (Lanier, 2013; [www.commodify.us](http://www.commodify.us)). In the context of data-driven medicine, the emphasis on increasing individual control over data has translated into attempts to develop better anonymization techniques and more fine-grained informed consent procedures (Kaye et al., 2015), as well as the configuration of patients as the rightful ‘owners’ of their own medical data (Kish and Topol, 2015).

However, scholars from different disciplines have begun questioning whether enhancing individual control over data is the most effective or desirable means of addressing the challenges raised by the increased digitalization of society in general, and the new power differentials it enables in particular. Various strands of social theory and feminist scholarship, for example, have critiqued the pervasive understanding of persons as autonomous individuals that underpins the notion of individual control over data, arguing for more relational and interdependent concepts of selfhood (e.g., Mackenzie and Stoljar, 2000). Others emphasize the social nature, not so much of personhood, but of data, which is never clearly individual but always shared to some extent, certainly in the case of medical and genetic data (Taylor, 2012). Some legal scholars, furthermore, doubt the legal feasibility of complete control, if not ownership, of data by one individual (Evans, 2016). And others remark the futility of monetization schemes for personal data as a means of redressing the colossal inequality between internet companies and users, which may inadvertently lead to the ‘proletarianization’ of users and the transformation of privacy into a luxury of those who can afford not to sell their data (Casilli, 2019). Perhaps most importantly, the emphasis on individual rights and values may result in a reframing of societal concerns as individual ones all the while undermining the political power of collectives.

Each of the articles, commentaries and interview that make up this special theme addresses the reconfiguration of existing relationships and the emergence of new power differentials that result from the expansion of the health data ecosystem. While they do this from different perspectives, they all share the same starting point: the understanding that increased individual control of data subjects is insufficient for anticipating the far-reaching risks and preventing the societal, if not individual, harms associated with this expansion. In light of this, they argue for new governance frameworks, technological infrastructures and narratives that are predicated on the *shared* responsibility of *multiple* stakeholders and *collective* decision-making and control. The contributions are the result of a two-day symposium funded by the Netherlands Organization for Science (NWO) held at the University of Maastricht in November 2017. The symposium brought together on the one hand humanities and social science scholars spanning law, sociology, Science and Technology Studies (STS), philosophy and critical data studies, and on the other, scholars and practitioners from the fields of computer science, bioinformatics and technology and health advocacy, to discuss how the expansion of the health data ecosystem disrupts existing norms and frameworks of data ethics and governance, and what kinds of re-thinking of ethics and governance this solicits in theory and in practice.

The commentaries by Brian Bot, Lara Mangravite and John Wilbanks, and by Bart Jacobs and Jean Popma both discuss the types of technical methods and arrangements that need to be developed to enable secure, responsible and equitable data sharing in the context of decentralized medical research. Both groups of authors are involved in the design and implementation of novel data management infrastructures. Bot et al., senior scientists at Sage Bionetworks – a pioneer in developing and operating platforms for distributed biomedical data collection and analysis – argue that the unique nature of digital medical data, which does not lend itself easily to ownership frameworks, requires a shift of focus from ownership to access and effective governance mechanisms that draw on a wide variety of tools including new types of (digital) consent, ‘model-to-data’ approaches, and data use agreements. Drawing on three examples of decentralized research that Sage has facilitated, they discuss the benefits and challenges of decentralization in terms of changing relationships between three important actors: research participants, primary and secondary researchers.

Jacobs and Popma also draw on their experience as designers of the data management infrastructure for an ongoing large cohort study on Parkinson's disease carried out by a university medical center in the Netherlands in partnership with Verily Life Sciences, an Alphabet subsidiary. Based on this work, they list a number of technical, organizational and legal conditions they believe should be met for responsible, data-intensive multi-stakeholder research collaborations: (1) Comprehensive but clear informed consent procedures; (2) Data governance that specifies the responsibilities and obligations of data controllers towards study participants that encompasses the entire life cycle of research data; (3) Legally binding data use agreements that ensure that the obligations of data controllers are met, that derived data become part of the study data repository to be shared under the specifications of the data governance framework and that non-compliance is sanctioned with revocation of data-sharing agreements; (4) Development of data protection mechanisms, such as the Polymorphic Encryption and Pseudonymisation technology that the authors discuss, to avoid leakage, hacks and unlawful combination of data.

A better understanding of the workings of data management infrastructures, and interdisciplinary collaboration with the computer scientists and bioinformaticians who are helping construct them, is not new to the fields of STS and data studies, and it should be more than ever fostered. In a filmed interview with José van Dijck on the recent book she has co-authored with Thomas Poell and Martijn de Waal, *Platform Society: Public Values in a Connective World* (2018), van Dijck and Sharon discuss the importance of grasping how the material functioning of internet platforms contributes to shaping a new political and social reality. According to van Dijck, the platform economy enables bigger players to elude classic taxonomies upon which much of our institutional and legal frameworks are predicated. Our governance systems traditionally depend on a division between infrastructures and sectors, explains van Dijck, but platforms introduce a new type of hybrid organization that blurs these categories, allowing them to bypass sectorial regulation. Sharon and van Dijck also discuss how the norms and values that are inscribed in the architecture of platforms collide with public values and 'de-bundle' collectives, how this plays out in the platformization of the health sector and who has the greatest responsibility in re-designing platform society in such a way that it would be anchored in public values.

In their commentary, Alessandro Blasimme, Effy Vayena and Ine Van Hoyweghen turn to a less commonly studied stakeholder: the private insurance sector. They scrutinize how the expansion of the health data

ecosystem is unsettling the position of private insurers, and the impact this may have on the willingness of people to participate in precision medicine initiatives. Such initiatives, like the American 'All of Us' program which they discuss, are predicated on the pooling together of vast amounts of data that are collected by participants themselves, via wearables and mobile devices. While this proliferation of citizen generation of medical data is a boon for research, it creates a new 'information asymmetry' between private insurers and those of their policy-holders who enroll in such research: policy-holders may know much more about their own health risks than their insurers. The authors explain that this will likely prompt insurers to claim access to these data about their prospective customers, which in turn will likely make people more reluctant to donate personal health data for precision medicine research. Here too, the authors argue for the need for new governance mechanisms that could mitigate this development by balancing the different interests of insurers, citizens and society as the beneficiary of scientific research. They suggest that a set of three principles – trustworthiness, openness, and evidence – should guide this.

Tuukka Lehtiniemi and Minna Ruckenstein propose a different approach, focusing on data activism as a productive means of challenging the power asymmetries of datafied societies. However, as they show, different 'social imaginaries', or different notions of desirable futures, underlie data activism, and they are not equally valuable. Based on their engagement as social scientists with MyData, a data activism initiative originating in Finland, the authors identify and disentangle two parallel social imaginaries and discuss their benefits and disadvantages. What they call the 'technological imaginary' is fed by practical and future-oriented aims, and favors technological solutions such as infrastructural interventions and monetization of personal data. What they call the 'socio-critical imaginary', conversely, questions the effectiveness of technological fixes to societal problems, and more importantly – it critically interrogates the assumption that increased individual control of data flows that frame the technological correctives of the 'technological imaginary' is a desirable or feasible aim. The authors call for a greater role for this socio-critical imaginary, with its sensitivity to the role of social structures and the political economy in shaping data futures, and its emphasis on the need for collective, rather than individual-centric, activism. However, they see room for improvement here too. The socio-critical imaginary would benefit from engaging with some characteristics of the technological imaginary, namely, with its practical orientation and infrastructural know-how. As they discuss, a productive synthesis of the two imaginaries is the best means of

making data activism more socially robust and responsible.

An example of the synthesis of imaginaries that Lehtiniemi and Ruckenstein argue for is the merging of infrastructural technologies with MyData principles about collective-centricity to produce data commons. Indeed, for those scholars and activists seeking to counter the power asymmetries that characterize our datafied present by foregrounding collective, rather than individual, control over data, data commons and cooperatives have become perhaps *the* preferred site of theoretical and practical resistance. But the commons framework also raises certain limitations. These are the focus of the contributions by Linnet Taylor and Nadya Purtova and Barbara Prainsack. Both pieces argue that the specific nature of digital data – namely what Prainsack calls their ‘multiplicity’, or the fact that they can be distributed in time and space – makes them substantially different from the physical resources that early commons scholarship studied, such as fisheries and farmlands (e.g., Ostrom, 1990). This means that there can be no simple transposition of the design principles for physical commons to data commons, something that data commons enthusiasts should ponder. This is not to say that data cannot be organized as commons, but that the original commons framework needs to be adjusted and expanded through rigorous analytical work. For Taylor and Purtova, this means more attention should be paid to the question of which stakeholders are affected by data practices and involving them in data governance on the one hand, and to developing governing institutions that would facilitate communication and trust between stakeholders and draw up (enforceable) rules for sustainable data use, on the other. Prainsack argues that a more systematic discussion of processes of inclusion and exclusion in commons is required. While Ostrom and her collaborators saw the possibility to exclude people from access, use or governance of commons as a condition of their sustainability, there is a tendency to view data commons, in line with egalitarian and democratizing narratives of the internet, as open access regimes from which no one can or should be excluded. This, Prainsack argues, is a recipe for appropriation by some and a reinforcement of rather than a counterweight to existing power asymmetries. Following Jodi Dean’s (1996) plea for more awareness and accountability in the inclusionary and exclusionary practices of solidarity, Prainsack calls for more awareness about what kinds of exclusion are appropriate for data commons that seek to redress power differentials, and the mechanisms required to prevent undue exclusion.

As these engagements with commons theory show, for collective-centred or commons-based approaches to be effective in redressing new power asymmetries

amongst old and new stakeholders in favour of collectives of data subjects – and for them to be more effective than individual-centred approaches – rigorous analytical attention must be paid. In her article, Tamar Sharon calls for a closer examination of the different conceptualizations of the common good that are at work in one specific area of the expansion of the health data ecosystem, what she calls the ‘Googlization of health research’, or the recent entrance of large consumer tech corporations into the domain of biomedical research and health. Using the framework of justification analysis (Boltanski and Thévenot, 2006), she identifies a plurality of conceptualizations of the common good that different actors mobilize to justify collaborating within these new multi-stakeholder research projects. This ethical pluralism, she argues, is not sufficiently accounted for in critical data studies and STS literature, yet identifying it is a necessary first step preceding the development of governance frameworks that seek to ensure the common good. Subsequently, these frameworks can combine repertoires of the common good in ways that secure that the civic conception, with its appeals to solidarity and social value, is central, but has been updated. Such combinations resonate with the synthesis of social imaginaries suggested by Lehtiniemi and Ruckenstein.

We hope that this special theme offers a productive – albeit far from comprehensive – overview of arguments for and examples of infrastructure, governance and ethics that are collective-centric in addressing the challenges posed by the datafication and expansion of the health ecosystem. We would like to thank *Big Data & Society* for the opportunity to align these different perspectives, the NWO for funding the symposium which originally brought them together and the reviewers who provided invaluable feedback.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID iD

Tamar Sharon  <https://orcid.org/0000-0003-0155-9220>

### References

- Andrejevic M (2014) The big data divide. *International Journal of Communication* 8: 1673–89.



- Boltanski L and Thévenot L (2006) *On Justification: Economies of Worth*. Princeton: Princeton University Press.
- boyd d and Crawford K (2012) Critical Questions for Big Data. *Information Communication & Society* 15(5): 662–79.
- Casilli A (2019) *En Attendant les Robots: Enquête sur le Travail du Clic*. Paris: Seuil.
- Dean J (1996) *Solidarity of Strangers*. Berkeley: University of California Press.
- Evans B (2016) Barbarians at the gate: Consumer-driven health data commons and the transformation of citizen science. *American Journal of Law & Medicine* 4: 1–34.
- Health Data Exploration Project (2014) *Personal Data for the Public Good: New Opportunities to Enrich Understanding of Individual and Population Health*. Calit2, UC Irvine and UC San Diego. Available at: [http://hdexplore.calit2.net/wp-content/uploads/2015/08/hdx\\_final\\_report\\_small.pdf](http://hdexplore.calit2.net/wp-content/uploads/2015/08/hdx_final_report_small.pdf) (accessed 13 May 2019).
- Kaye J, Whitley E, Lund D, et al. (2015) Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics* 23(2): 141–146.
- Kish L and Topol E (2015) Unpatients – Why patients should own their medical data. *Nature Biotechnology* 33(9): 921–924.
- Lanier J (2013) *Who Owns the Future?* London: Penguin Books.
- Lucivero F and Prainsack B (2015) The lifestylisation of healthcare? “Consumer genomics” and mobile health as technologies for healthy lifestyle’. *Applied and Translational Genomics* 4: 44–49.
- Mackenzie C and Stoljar N (2000) *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. Oxford: Oxford University Press.
- Ostrom E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Purtova N (2015) The illusion of personal data as no one’s property. *Law, Innovation and Technology* 7(1): 83–111.
- Sharon T (2016) The Googlization of health research: from disruptive innovation to disruptive ethics. *Personalized Medicine*. DOI: 10.2217/pme-2016-0057. Available at: [https://www.futuremedicine.com/doi/abs/10.2217/pme-2016-0057?rfr\\_dat=cr\\_pub%3Dpubmed&url\\_ver=Z39.88-2003&rfr\\_id=ori%3Arid%3Acrossref.org&journalCode=pme](https://www.futuremedicine.com/doi/abs/10.2217/pme-2016-0057?rfr_dat=cr_pub%3Dpubmed&url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&journalCode=pme).
- Shen H (2015) Smartphones set to boost large-scale health studies. *Nature*. DOI:10.1038/nature.2015.17083. Available at: <https://www.nature.com/news/smartphones-set-to-boost-large-scale-health-studies-1.17083>.
- Taylor L (2017) What is data justice? *Big Data & Society* 4(2): 1–14.
- Taylor M (2012) *Genetic Data and the Law: A Critical Perspective on Privacy Protection*. Cambridge: Cambridge University Press.
- Topol E (2015) *The Patient Will See You Now: The Future of Medicine Is in Your Hands*. New York: Basic Books.
- van Dijk J, Poell T and de Waal M (2018) *The Platform Society: Public Values in a Connective World*. Oxford: Oxford University Press.
- Weber GM, Mandl KD and Kohane IS (2014) Finding the missing link for big biomedical data. *JAMA* 311(24): 2479–2480.
- Zuboff S (2015) Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30(1): 75–89.