swiss medical weekly

Viewpoint | Published 06 January 2023 | doi:10.57187/smw.2023.40019
Cite this as: Swiss Med Wkly. 2023;153:40019

Medical data sharing and privacy: a false dichotomy?

Marcello Ienca[a,b]

[a] Intelligent Systems Group, College of Humanities, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[b] School of Medicine, Institute of History and Ethics in Medicine, Technical University of Munich, Munich, Germany

Introduction

The relationship between medical data sharing and privacy has long been a focus of debate in areas such as medical ethics, medical informatics and health policy. However, this debate has often remained within the sphere of the speculative, quite disconnected from empirical data. A recent cross-sectional survey from Switzerland [1] published in the *Swiss Medical Weekly* provides useful insights on the nexus between data sharing and privacy, and corroborates previous empirical evidence on this topic.

Data-driven medicine and privacy

In the last three decades, two main socio-technological trends have emerged that have irreversibly changed both clinical medicine and biomedical research: data-driven medicine and privacy protection.

The first trend concerns the increasing availability of data in digital format (thereby called *"digital data"*), including data related to human health (hereafter called *"health data"*). Thanks to *inter alia* the digitization of medical records [2], progress in data storage capacity and the development of so-called "digital phenotyping" technologies [3], it is now possible to acquire and share much larger and more heterogeneous volumes of health data than in the past. In certain cases, the volume and heterogeneity of the data exceed the processing capabilities of human analysts. Therefore, the analytic velocity of automatic computational systems (generally based on so-called *machine learning*) is needed. It has become common place to refer to those cases as "big data" [4]. Originating in the astronomical sciences, the socio-technical phenomenon of big data now also largely characterizes the biomedical and healthcare fields, especially in areas such as genomics, epidemiology, medical imaging and healthcare management [5]. The benefits of increased data availability and sharing are many but can be classified into three main categories:

– *Precision medicine:* Rich data inherent to the specific characteristics of individual patients can enable the development of an individualised, and thus more precise, approach for each patient.
– *Prevention and early diagnostics:* Large volumes of data can be analysed using machine learning algorithms to detect abnormalities more quickly than with traditional techniques, thus enabling more efficient preventive strategies and diagnostic techniques.
– *Monitoring:* The possibility of acquiring continuous (i.e. not temporally discrete) data on both individual subjects and groups of people allows for more efficient monitoring of both individual and public health.

The second trend is not technological but ethical and sociological in nature. It concerns the growing interest of the general population in protecting their personal data and, more broadly, in protecting their informational privacy. Informational privacy can be defined as the right and ability of an individual or group to seclude themselves or information about themselves. According to some, privacy is a necessary component of personal autonomy, as it constitutes a necessary requirement for people to express themselves autonomously and selectively [6, 7].

Privacy is not a new concept. In the ancient world, Aristotle distinguished the public sphere of the *polis,* or the city-state, from the private sphere of the *oikos,* or the private sphere encompassing the family, the family's property, and the house [8]. However, this characterisation drew a rudimentary separation between spaces, not between information flows. In the late 19th century, legal scholars Warren and Brandeis developed a more information-focused understanding of privacy, and recognised a proto-right to privacy which they defined as the right "to be let alone" [9]. This account laid the foundations of what we now refer to as "informational privacy" (also called "information privacy") [10, 11].

Although the idea of privacy has persisted throughout much of the history of Western thought, it is plausible that the recently increased interest in privacy of both the public and legislators has been induced precisely by the increasing, sometimes pervasive, availability of digital data in the last three decades. In particular, the aggressive, unbridled acquisition of digital data by private entities in fields such as telecommunications and social media has raised concerns about so-called "surveillance capitalism" [12] and prompted philosophers and scientists to emphasise the ethical-legal importance of the right to privacy [13]. This trend has been exacerbated by media-amplified data breaches that have severely undermined public trust in the free acquisition and sharing of data [14, 15]. Some of these breaches have involved health data [16].

Correspondence:
Dr Marcello Ienca
School of Medicine
Institute of History and Ethics in Medicine
Technical University of Munich
Ismaninger Straße 22
DE-81675 Munich
marcello.ienca[at]epfl.ch

Medical data sharing and privacy: a complex relationship

In light of this historical excursus, a fundamental question arises: What relationship exists between data sharing in medicine and privacy?

The default answer to this question is that increasing data acquisition and sharing logically and ethically conflicts with the protection of privacy. To put it simply: The more data are acquired and shared, the more privacy is eroded.

This thesis was put forward by the famous epidemiologist Gilbert W. Beebe, who wrote back in 1983 that in order *"to cope with the increasing demands of our society for prevention, treatment, and compensation, we need more precise information on health hazard"* which, in turn, *"will require better planning and integration of existing information systems, additional funds, and some trifling sacrifice of personal privacy"* [17].

In the context of individual health by Choudhury et al. have argued that the "*researchers' willingness to share data can also be constrained by concerns for the privacy of the human research participants who are the data sources*" [18].

In the field of medicine, the default position on the relationship between data sharing and privacy requires that if individuals or groups want to protect their informational privacy, then they should refrain from sharing their health data for biomedical research, or at least minimise such activities. This hypothesis is empirically testable. Should it be true, we should observe from survey studies that people who have a greater interest in protecting their privacy will have a lower willingness to share their medical data.

Scrutinising the empirical evidence

However, recent empirical studies do not seem to support this hypothesis. In fact, some studies seem to falsify it (in the Popperian sense [19]). For example, several studies have shown that data subjects who strongly believe in their right to privacy are nonetheless willing to share their data provided that some conditions are met. These conditions include that (1) data subjects feel empowered to make informed and unimpeded decisions, (2) data management policies are transparent, (3) adequate technical safeguards are in place, and (4) data processors are perceived as trustworthy.

For example, the experience of the International Cancer Genome Consortium (ICGC) has shown that a controlled access mechanism that contains privacy safeguards and preserves the autonomy of data subjects is likely "to reconcile open data sharing with privacy concerns" [20]. Similarly, Leon et al. have shown that whenever clear data-retention and scope-of-use policies were in place, participants were more willing to allow data collection [21]. Using questionnaire methods, Caine and Hanania have shown that patients have a pronounced preference for sharing data under granular privacy regimes that allow them to have control over which information will be shared and with whom [22].

Finally, the afore mentioned cross-sectional survey by Pletscher et al. published in this journal provides further evidence on the nexus between data sharing and privacy. The survey results show that although privacy and data protection concerns are very common among the Swiss population (74%), the large majority (71%) of respondents (with peaks of 81% among people with chronic diseases) reported that they are nevertheless willing to share their data for medical research [1] provided that a number of conditions are met. These conditions largely coincide with those identified by previous studies as they include data anonymisation, clear public health benefit, trustworthiness of the data processing institutions (which appeared higher for hospitals and universities and lower for pharmaceutical and insurance companies), and the explicit indication of the purposes for which the data will be used.

These results also corroborate the results of a previous large-scale Swiss survey on public willingness to participate in personalised health research and biobanking by Brall et al. This latter survey showed that certain types of health data such as questionnaires about health status and blood samples would be willingly shared by more than 80% of the Swiss population [23].

Conclusion and room for future research

In light of these studies, the time seems ripe to overcome the anecdotal and empirically unjustified dichotomy between data sharing and privacy. Indeed, considering privacy and data sharing as mutually exclusive seems to be an informal logical fallacy known as a false dichotomy (or false dilemma).

According to the available empirical evidence, data subjects do not seem to choose between either sharing their data or protecting their privacy. On the contrary, they seem to make decisions about the sharing of their data based on privacy and personal autonomy considerations. In other words, it is not a question of *choosing between* privacy and data sharing, but of *determining which* privacy (and more generally also security, transparency and trustworthiness) requirements are necessary to ensure an efficient degree of data sharing in medical research. Studies such as those mentioned above are providing us with useful information in this regard. Now we must convert this information into practice.

However, future research isneeded to elucidate how different types of health data may generate different degrees of willingness to share data. In fact, a recurring limitation of survey studies in this area is that they often treat health data as if they were an uniform category of data. Yet, it is likely that different data types generate different expectations of privacy and different degrees of willingness to engage in data sharing. This hypothesis is partly confirmed by Brall et al [23] as they have shown that >80% of Swiss respondents are willing to share blood sample data or data from health questionnaires. However, less than 40% are willing to share data from health apps. By treating "health data" as a monolithic category, there is a risk of neglecting these possible variations between data types and data sharing attitudes. Similarly, both Pletscher et al. [1] and Brall et al. [23] also showed that a complex matrix of factors is at stake when it comes to predicting data sharing willingness. These include considerations related to the utility of results, data governance and management mechanisms, as well as personal motivations and concerns. As Pletscher et al. have shown, data anonymisation appears to be a crucial

factor. Future research should explore this matrix of factors and identify predictors of willingness to share data specific to each data category and socio-cultural context.

References

1. Pletscher F, Mändli Lerch K, Glinz D. Willingness to share anonymised routinely collected clinical health data in Switzerland: a cross-sectional survey. Swiss Med Wkly. 2022 Jun;152:w30182. http://dx.doi.org/10.4414/SMW.2022.w30182.
2. Miller RH, Sim I. Physicians' use of electronic medical records: barriers and solutions. Health Aff (Millwood). 2004 Mar-Apr;23(2):116–26. http://dx.doi.org/10.1377/hlthaff.23.2.116.
3. Insel TR. Digital phenotyping: technology for a new science of behavior. JAMA. 2017 Oct;318(13):1215–6. http://dx.doi.org/10.1001/jama.2017.11295.
4. Marx V. Biology: the big challenges of big data. Nature. 2013 Jun;498(7453):255–60. http://dx.doi.org/10.1038/498255a.
5. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: A scoping review. PLoS One. 2018 Oct;13(10):e0204937. http://dx.doi.org/10.1371/journal.pone.0204937.
6. Childress JF. The place of autonomy in bioethics. Hastings Cent Rep. 1990 Jan-Feb;20(1):12–7. http://dx.doi.org/10.2307/3562967.
7. Childress JF, Beauchamp TL. Principles of biomedical ethics. Oxford University Press Oxford; 1994.
8. Barker E, Stalley RF. (Oxford: Oxford University Press, 1995).
9. Warren S, Brandeis L. The right to privacy. Harv Law Rev. 1890;15.
10. Floridi L. Four challenges for a theory of informational privacy. Ethics Inf Technol. 2006;8(3):109–19. http://dx.doi.org/10.1007/s10676-006-9121-3.
11. Tavani HT. Informational privacy, data mining, and the internet. Ethics Inf Technol. 1999;1(2):137–45. http://dx.doi.org/10.1023/A:1010063528863.
12. Zuboff S. Big other: surveillance capitalism and the prospects of an information civilization. J Inf Technol. 2015;30(1):75–89. http://dx.doi.org/10.1057/jit.2015.5.
13. Véliz C. Privacy is power. Melville House; 2021.
14. Isaak J, Hanna MJ. User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer. 2018;51(8):56–9. http://dx.doi.org/10.1109/MC.2018.3191268.
15. Ienca M, Vayena E. Cambridge analytica and online manipulation. Sci Am. 2018;30.
16. Wikina SB. What caused the breach? An examination of use of information technology and health data breaches. Perspect Health Inf Manag. 2014 Oct;11:1h.
17. Beebe GW. Vol. 73 245-246 (American Public Health Association, 1983).
18. Choudhury S, Fishman JR, McGowan ML, Juengst ET. Big data, open science and the brain: lessons learned from genomics. Front Hum Neurosci. 2014 May;8:239. http://dx.doi.org/10.3389/fnhum.2014.00239.
19 Popper, K. R. Science as falsification. Conjectures and refutations 1, 33-39 (1963).
20. Joly Y, Dove ES, Knoppers BM, Bobrow M, Chalmers D. Data sharing in the post-genomic world: the experience of the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO). PLOS Comput Biol. 2012;8(7):e1002549. http://dx.doi.org/10.1371/journal.pcbi.1002549.
21. Leon PG, et al. in Proceedings of the ninth symposium on usable privacy and security. 1-12.
22. Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. J Am Med Inform Assoc. 2013 Jan;20(1):7–15. http://dx.doi.org/10.1136/amiajnl-2012-001023.
23. Brall C, Berlin C, Zwahlen M, Ormond KE, Egger M, Vayena E. Public willingness to participate in personalized health research and biobanking: A large-scale Swiss survey. PLoS One. 2021 Apr;16(4):e0249141. http://dx.doi.org/10.1371/journal.pone.0249141.