IIC (2021) 52:821–824
https://doi.org/10.1007/s40319-021-01067-5

EDITORIAL

Who Controls Covid-Related Medical Data? Copyright and Personal Data

Michael Birnhack

Accepted: 6 April 2021 / Published online: 17 May 2021

In January 2021, Pfizer, who together with BioNTech developed a vaccine for Covid-19, and the Israeli Ministry of Health (MoH) signed an agreement entitled the *Real-World Epidemiological Evidence Collaboration*. Per the agreement, Pfizer would provide vaccines for Israel's entire eligible population and receive an undisclosed payment and data regarding the vaccine's safety and efficacy. Israel thus served as a *de facto* laboratory, with the world closely following. The agreement's declared goal was to track pandemic data in a real-world context. The MoH would share "aggregate project data" (Sec. 3), meaning "any de-identified data" (Sec. 1.8). "Project data" are owned by the MoH or the Israeli Health Maintenance Organizations (HMO) (Sec. 8.1).[1] The data are about millions of people from a variety of sources and are used as they are collected. This is big medical data.

Who controls big medical data of such utmost significance? The agreement highlights the interrelationship between two modes of protecting data: data protection law and copyright law. The former provides legal protection to data subjects and limits the data controller, but generally speaking, allows the data to be processed for the benefit of public health; the latter awards the databases' controller with rights regarding the dataset, a control which may hinder others' access to highly important data. This editorial unpacks this relationship.

[1] *See* https://govextra.gov.il/media/30806/11221-moh-pfizer-collaboration-agreement-redacted.pdf.

M. Birnhack (✉)
Associate Dean for Research, Professor of Law, Faculty of Law, Tel Aviv University, Tel Aviv, Israel
e-mail: birnhack@tauex.tau.ac.il

1 Data Protection and Big Data

When the data are about a living person, we summon data protection law. Under the European Union's General Data Protection Regulation (GDPR), personal data – defined as "any information relating to an identified or identifiable natural person" – deserves legal protection.[2] The ideal of data protection law is that only the data subject may determine whether to share the data, with whom, when, how, and under which conditions. This is the idea of (informational) privacy as control. The law permits processing of personal data with the subject's free and informed consent or in certain other cases.[3] For scientific research, which includes medical data, the GDPR offers an explicit derogation, subject to some safeguards (Art. 89).

Some scholars have argued in favor of treating personal data as property, but their arguments have been refuted: Treating personal data as property would lead to more, not less, human commodification and would play into the hands of data giants rather than data subjects.[4] Data about oneself may be precious to the individual and serve as a shield against actions such as discrimination or harassment. Controlling one's data facilitates autonomy and protects human dignity. However, the data is also valuable to others: the state, employers, insurance companies, and data giants. Digital data mongers are often interested in big data rather than in an individual's data; they analyze the dataset and ultimately target individuals.[5]

Access to big medical data enables the study of pre-existing data rather than of people in clinical trials. It is safer and cheaper. Data analytics enables searching for unknown correlations, which leads to searching for causations. Instead of being limited to a sample population, the study can cover more people with more diverse relevant backgrounds. Notwithstanding the benefits of big medical data studies, they pose a substantial challenge to data protection law: The data were collected in an identifying mode as a byproduct of medical treatment, and are then de-identified. The researchers often wish to have as much data as possible about any individual, such as which neighborhood they live in, as one may be more polluted than the other; or perhaps ethnicity matters, as it reflects genetics? In other words, during collection, data subjects are identified; during processing, data subjects are identifiable; and at publication, the data are aggregate and statistical. A straightforward application of data protection law

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1 (GDPR), Art. 4(1).

[3] GDPR, Art. 6(1) offers six legal bases for such lawful processing.

[4] Lawrence Lessig (1999) *Code and Other Laws of Cyberspace*, pp. 159–162. For criticism, *see* Julie E. Cohen (2000) "Examined Lives: Informational Privacy and the Subject as an Object", 52 Stan. L. Rev. p. 1373.

[5] Shoshana Zuboff (2019) *The Age of Surveillance Capitalism*: *The Fight for a Human Future at the New Frontier of Power*. Big data pose their own challenges to data protection law. *See* Omer Tene and Jules Polonetsky (2013) "Big Data for All: Privacy and User Control in the Age of Analytics", 11 NW. J. Tech. & Intell. Prop. p. 239.

would require that data collection and processing be permitted only with the subject's consent.[6]

If the data are truly anonymized, then data subjects are *de facto* protected. However, with the advance of de-anonymization and re-identification techniques, anonymity is fragile.[7] When the data subjects are identified, they have rights, and the data controller and data processor are subject to various obligations.

2 Copyright and Big Data

Copyright law places a "No Entry" sign and refuses to protect raw data.[8] Copyright law's refusal to protect raw data is a well-entrenched legal maxim for several reasons. Firstly, a fact about the world was revealed, explained, and articulated by an author but not created by her. Secondly, data comprise raw material for further creativity, reflecting the notion of human knowledge being a collective human endeavor, echoing the idea of progress.[9] Thirdly, the free circulation of information has an important democratic role. In short, in the famous words of Justice Brandeis, referring to "knowledge, truths ascertained, conceptions, and ideas," and extended to data – it should be "free as the air to common use."[10]

Turning from single datum to big data, copyright law offers thin protection for structured, non-trivial compilations of data (i.e. selected and organized data), but not for the underlying data.[11] However, when the dataset is unstructured (i.e. without pre-defined selection criteria and no particular arrangement, and with each datum tagged), there is not much to protect under copyright law. The data aggregators achieve control of their datasets through other means, such as technological locks and trade secrets law.

Importantly, copyright ownership of big medical data may hinder others' access. The importance of the medical datasets, especially in times of a crisis, cannot be exaggerated. The wise use of the data can decidedly save many lives. Such control may be a copyright ideal for some but a drawback for all.

[6] Michael Birnhack (2019) "A Process-Based Approach to Informational Privacy and the Case of Big Medical Data", 20 Theoretical Inq. L. p. 257.

[7] Paul Ohm (2010) "Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization", 57 UCLA L. Rev. p. 1701.

[8] Berne Convention for the Protection of Literary and Artistic Works, Art. 2(8).

[9] Michael Birnhack (2001) "The Idea of Progress in Copyright Law", 1 Buffalo Intell. Prop. L.J. p. 3

[10] *International News Service v. Associated Press*, 248 U.S. 215, 250 (1918).

[11] Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS), 15 April 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299 (1994), Art. 10(2). TRIPS does not cite the originality requirement, but this is taken to be a standard legal norm, despite its different interpretations. *See, e.g.*, *Feist Publications, Inc. v. Rural Telephone Service Company, Inc.*, 499 U.S. 340 (1991); *CCH Canadian Ltd. v. Law Society of Upper Canada*, 2004 SCC 13, [2004] 1 S.C.R. 339. Databases may also be protected under the *sui generis* European Directive. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

3 Interaction

Combining the two bodies of law regarding big medical data results in four primary situations:

(a) Structured, anonymous datasets: The controller enjoys copyright, and data subjects enjoy *de facto* privacy protection.
(b) Unstructured, anonymous datasets: The controller does not enjoy copyright protection and may revert to other means, such as trade secrets law. Data subjects enjoy *de facto* privacy.
Both (a) and (b) carry the risk of de-anonymization, especially if the controller shares the dataset with others.
(c) Structured, identifying datasets: The controller enjoys copyright protection but is subject to various obligations to protect data subjects' privacy. Sharing the dataset would pose a high privacy risk. Data protection law reinforces copyright.
(d) Unstructured, identifying datasets: The controller does not enjoy copyright protection and is under an obligation to protect data subjects' privacy, pushing the controller to seek other restrictive practices of not sharing the data.

Returning to the Israeli context, according to the National Health Insurance Act, HMOs hold detailed medical data about the entire population, including vaccinations and their side effects. The HMOs can be classified under situation (c), with additional regulatory duties of confidentiality. The two largest HMOs conduct medical research also under situation (d). When the HMOs transfer aggregate and statistical data to the MoH, which, in turn, transfers it to Pfizer, we shift to situation (a), thus reducing privacy risks.

Combining copyright law and the obligations imposed by data protection law pushes the parties to protect the data under both copyright law and additional layers of protection, such as trade secrets law. This result means that other parties may have access to outcomes but not to raw data. To facilitate broad access to crucial data during a global health crisis, we need to address both bodies of law in an integrated manner.