Health data ownership: status, importance, open problems

Final Paper for Data Science Society

Ilaria Battiston - 03723403

Winter Semester 2021

Contents

1 Introduction to data ownership **3**

2 Personal data ownership **3**
2.1 The healthcare scenario 4

3 Uses of personal data and morality **4**

4 Political and legal issues **5**

5 Open source and open data **7**

6 Conclusion **8**

1 Introduction to data ownership

The amount of data in the world nowadays is substantially increasing: many of the personal and non-personal aspects of people's everyday activities are aggregated and stored as data by both companies and governments.

Data ownership, as stated by [1], is defined *as the act of having complete control on a set of data*, including legal rights, distribution policy and license. In other words, the party having ownership has the ability to create, edit, modify and restrict access to data, claiming copyright on it.

Ownership, therefore, implies both **possession** and **responsibility** for information: having control naturally allows to decide whether data should be shared. On the other side, someone who is not the owner should not have access.

It might seem trivial that an individual producing data should also be its owner: in the fields of art, industry or science, the figure of creator is distinct from the reader, and ownership policies are generally strict.

For instance, enterprises producing big data regarding their products have full ownership and heavy restrictions to be imposed to whoever uses this information: employees and research teams are subject to non-disclosure agreements, and results are only published in the form of aggregated information.

2 Personal data ownership

The concept stated above tends to lose part of its simplicity when dealing with personal data, resulting in a somewhat gray area.

Typically, organizations can capture different personal data in multiple ways, as defined by [10]:

- Data can be **volunteered** by individuals when they explicitly share information about themselves through electronic and social media;
- Data can be **obtained** while providing a service;
- Observed data are **captured** by recording users' activities;
- Data can also be **inferred** based on analysis of trends and behaviors.

So, in all this chaos of data generation and delegated access, where does true ownership lie? There are some situation in which this is not clear, or is often misunderstood and

misused. This opens a lot of questions: is ownership the same as private property? Who could be entitled to claim the data?

2.1 The healthcare scenario

A trivial real-life scenario in which problems arise is healthcare. Health data does not only concern medical information: it involves anything related to person's life affecting medical decisions, such as *salary*, *geographical area*, *profession* and such.

Individuals may be used to expecting some kind of tracking, but without realizing that their healthcare data is used in such wide ways for the profit of others.

First of all, there are **no clear laws** on data ownership: a patient's privacy is protected, but this does not imply a patient always has the rights to delete data or give consensus to its use by the organization providing the healthcare service.

Second, health data has **multiple consumers**: pharmaceutical companies, hospitals, governments, data brokers, researches and analytical teams: this can often lead to a sort of *dehumanization* of individuals, giving more importance to profit and business decisions rather than privacy.

The doctor-patient relationship should be noble and not influenced by business factors, however it risks to be conditioned by the economy and the purpose of making profit. To state a real-life example, a major cancer researcher in the Sloan Kettering Cancer Center failed to disclose corporate financial ties in major research journals, incurring in conflicts of interests and breaking ethics [3].

Furthermore, a startup in the field of artificial intelligence whose stocks were owned by Memorial Sloan Kettering had exclusive rights to deal with data of 25 millions of patients while receiving funding by public authorities. This work was done over 60 years, raising plenty of concerns about the compliance with morals and laws.

In such cases, patients have absolutely zero ownership on their data, which was ultimately used for monetary purposes, and there are plenty of similar cases. Even the CMO of Moderna, one of the companies producing the COVID-19 vaccine, has sold almost 70 millions of dollars in stocks of his own firm to privately make profit [5].

3 Uses of personal data and morality

Since the sources of data are endless and it can be generated in real time, user datasets are well suitable for artificial intelligence and machine learning.

This arises the biggest argument against data ownership by individuals: ownership should have **control** as a direct consequence, but health data generated about the patient by a provider should be co-owned by both parties. On the other hand, data generated only by the patient should imply full ownership with rights to possess, share, sell or destroy it.

Patients' *fundamental right to protection of their health data* is an important issue in the healthcare context: in this case, even though the subject of information is a patient, there is an additional figure of **controller**, the entity (whether public or private) which collect and process personal data and determine the purpose and means for processing the data.

Even though data protection is considered a fundamental right, it is often given as a counterargument that the processing of health data is essential for the good functioning of healthcare services and improve public health [7].

Patients organizations are also gathering and using patients' data in their advocacy or research activities, so being able to use patients' personal data is sometimes important to advance research, healthcare practices or patients' rights.

Implicit in the previous argument is the assumption that patients would want to benefit from their clinical data. In general, this is not supported by empirical research, which consistently shows that patients are uncomfortable with the commercialization of clinical health data, in particular when the purpose is mere profit [6].

For the reasons stated above, it is important that healthcare services providers are aware of the rights of patients in this area and engage in order to ensure that the patients' perspective on data sharing.

However, data ownership is not something most people are aware of: consent to information processing is usually as simple as ticking a box, whose meaning is easily lost in the process of compiling several forms and questionnaires.

Patients, therefore, may be *overestimating* their ability to protect the data: even if they were required to consent to all secondary uses of their clinical data, they are unlikely to be in a position to fully assess these features and therefore make informed decisions.

4 Political and legal issues

European laws state that health data can be processed if it is in the patient's vital interest or for assessing and diagnosing diseases; the meaning of "processed" is yet not clearly stated. First of all, consent is not needed as long as parties are bound by

professional secrecy, which means professionals and healthcare institutions can easily communicate with one another.

This is advertised as a positive consequence of new regulations, but is it necessarily so? Is it worth to sacrifice part of the patients' privacy to facilitate the job of doctors?

Data for research purposes is also exempt from consent, as long as there is some form of *pseudoanonymisation*, i. e. the encoding of sensitive data in a way that it becomes almost impossible to identify the owner.

This can be easily perceived as hashing name and surname of patients, a theoretically flawless execution. However, it is simple to imagine how combining information such as GPS data with doctor location and past appointment times would easily enable to pinpoint an individual.

As stated by [4], even given the de-identification of patient health information, this does not prohibit the sale of data, require that health data is only used for patients' benefits, or protect it outside of the health system. In fact, consumer companies such as DNA testing and analysis services are not required to de-identify data if they sell it to data brokers.

Patients should also be able to access their own data and withdraw consent, yet is this really intelligible? Are individuals aware of the reasons why they could withdraw consent, and of the steps to take in order for this to be possible? Are data accessibility practices clearly stated and explained even to whom who do not have technology knowledge?

European regulations **do not explicitly state that access must be provided for free**, and allow controllers to charge a fee for administrative costs in case the subject asks more than once.

Transparency of information does not include any of the subjects stated above, in fact it just relates to the obligation of providing purpose, location, period and identity of any entity using healthcare data.

There is an exemption, in the case of research, if it proves to be a "disproportionate burden" to provide this information to data subjects. When data subject's requests for information are "unfounded" or repetitive, controllers can charge you to provide this information. Should people really be charged if they want to be aware of how their personal data is being used?

GDPR does not outline ownership either, and is very weak in general regarding this issue: according to it, *natural persons should have control of their own personal data.* The Right of Access recital states that *a data subject* ***should*** *have the right of access*

to personal data [...], including their health.

Overall, the European Union still does not have a clear position on data ownership: its laws are merely a series of data-related rights, including intellectual property rights, trade secrets, data protection rights, and other emerging forms of protections. However, this mainly concerns privacy, and does not seem sufficient to address the multiple facets of the issue.

In countries outside of the European Union, where GDPR does not apply, there is also not much agreement on data ownership, making it even more justifiable to always ask for the consent of the patient [9].

5 Open source and open data

Open source has grown into a way of participating with many others that asks for transparency, community-based collaboration, and meritocracy: these principles are very similar to problems raised by data ownership, hence the solution to them might lie among the set of ideas open source is based on.

Open data is getting more and more popular: applications allowing to manage electronic health records are starting to become open source, which means ultimately an individual should earn health autonomy by knowing who this data is shared with and most importantly be able to delete it.

Furthermore, there are open source friendly licenses and agreements clearly stating the terms of sharing medical knowledge.

Using an open data framework would allow full transparency on transactional care, and rights of viewing, accessing, sharing and destroying health information.

Furthermore, establishing open source policies would allow to maintain and evolve a single, national target for *health data standardization*, clearly defining its stakeholders and avoiding to give exclusive rights to one single party.

The only country using open national digital healthcare is Estonia: it implemented a policy aimed to increase transparency and easily provide access to information, decreasing the number of requests and facilitating data management.

The Open Data Portal provides a single point of access for general public to unrestricted public sector data with the permission to re-use and redistribute such data for both commercial and non-commercial purposes.

More than 70 data set can be found on the Estonian Open Government Data Portal, of which Tallinn is the top publisher [12].

6 Conclusion

Healthcare should be **fully open**, both in the infrastructure and with a clear definition of its ownership.

It is of extreme importance that governments play a more active role in restructuring their existing policy framework to protect personal data, inevitably needing to redesign and enforce data protection privacy laws and legislations.

This will require establishing policies at both the national and international level: entities will need to open up dialogue to establish comprehensive data protection and privacy laws that could be implemented globally.

This should be paired with a clearly articulated set of standards and procedures, which should go hand-in-hand with the open source philosophy and define responsibilities regarding procedures.

Patients need ownership over health data, which is often the most sensitive information about their lives, and requires an obvious need of authority over its access and use. Furthermore, it must be easily recognizable who is using and viewing this information.

Health data concerns everyone and everyone should demand ownership and rights, because this is the only way in which individuals can earn autonomy.

References

[1] "*Data Ownership*", Techopedia, no publication date. Accessed on: Mar 10, 2021. [Online]

[2] F. Banterle, *Data Ownership in the Data Economy: A European Dilemma*, SSRN Electronic Journal, 2018

[3] C. Ornstein, K. Thomas, *Memorial Sloan Kettering Leaders Violated Conflict-of-Interest Rules, Report Finds*, The New York Times, Apr 2019. Accessed on: Mar 10, 2021. [Online]

[4] S. Lee, *Who uses my health data?*, GoInvo, no publication date. Accessed on: Mar 8, 2021. [Online]

[5] S. Nagarajan, *Moderna's CEO sold nearly $2 million of his stock ahead of the company's emergency use vaccine filing*, Business Insider, Nov 2020. Accessed on: Mar 10, 2021. [Online]

[6] A. Ballantyne, *How should we think about clinical data ownership?*, Journal of Medical Ethics, 2020

[7] European Patient's Forum, *The new EU Regulation on the protection of personal data: what does it mean for patients?*, EPF, no publication date. Accessed on: Mar 9, 2021. [Online]

[8] B. Van Asbroeck, *Big Data & Issues & Opportunities: Data Ownership*, Bird&Bird, Mar 2019. Accessed on: Mar 10, 2021. [Online]

[9] T. Hulsen, *Data Ownership in Healthcare*, International Journal of Environmental Research and Public Health, 2020

[10] A. Al-Khouri, *Data Ownership: Who Owns 'My Data'?*, International Journal of Management & Information Technology, 2012

[11] GoInvo, *Open Source Healthcare*, Open Source Healthcare, no publication date. Accessed on: Mar 13, 2021. [Online]

[12] Tallinn, *Regions and Cities contribute the most of open data in Estonia – Tallinn is the top publisher overall*, Tallinn, May 2020. Accessed on: Mar 11, 2021. [Online]