Sassan Sangsari, MD & Dr. Natalie Rotermund
Artificial Intelligence Center Hamburg (ARIC) e.V.:

THE RISE OF HEALTH DATA USAGE

A CRITICAL ANALYSIS OF THE TENSION BETWEEN PROGRESS AND PRIVACY IN THE DIGITAL HEALTH REVOLUTION

aric

ABSTRACT

February 12th 2024 – The use of health data for research and product innovation promises a new era of medical progress and personalized treatment. How can data be used for the desired purposes while at the same time protecting data from unwanted uses? On the one hand, there are hopes of groundbreaking medical breakthroughs and economic opportunities. On the other hand is the right to informational self-determination, which emphasizes the privacy and protection of individuals' personal data. Between these two poles, there is an intense debate about the right balance between progress and privacy.

With the General Data Protection Regulation (GDPR), which came into force in 2018, the European Union has set the highest international legal standard for the protection of natural persons when handling their data. Even though the GDPR was introduced equally in every EU member state, it has historically been interpreted and applied most strictly in Germany.

A GDPR-compliant use of health data in Germany could therefore rightly be seen as a particular challenge, which, if successful, would also serve as a global role model.

This article takes a critical look at the challenges and debates surrounding the use of health data against the backdrop of the German approach to the GDPR. Central considerations are highlighted in two opposing poles for easier classification: the tensions between privacy vs. research, research vs. marketing, pseudonymization vs. anonymization, opt-in vs. opt-out, data provision vs. data use, as well as government IT projects vs. market-based solutions.

INFORMATIONAL SELF-DETERMINATION VS. RESEARCH & INNOVATION

Two fundamental interests collide in the debate on the use of health data: the informational self-determination of patients and the hope of progress through research in the healthcare sector. Society is faced with an ethical challenge in which a balance must be found between individual rights and the common good. In the context of medical research, real patient data is the main ingredient for the development of artificial intelligence (AI) and other groundbreaking technologies that promise enormous medical progress. At the same time, however, the use of such data has the potential to jeopardize atients' informational self-determination.

The importance of informational self-determination can hardly be overestimated. It enables people to protect their thoughts, wishes and personal information from others. The ability to be intimate and create special relationships depends on the ability to be more transparent to some people than to others. As the contemporary philosopher Moshe Halbertal aptly points out, the distinction between inside and outside would be obsolete if a person's thoughts were visible on their forehead. Privacy therefore p

rotects everyone's individual identity and allows them to define themselves as unique individuals. In this context, medical data is often particularly sensitive, as it contains information on mental and physical health, the extensive information of which has the potential to turn patients into transparent individuals.

In the context of medical research, this interest in self-determination comes up against the central challenge and ethical obligation of society to provide medical progress and the best possible care for its citizens. Real patient data containing accurate information about diagnoses, treatments and medical histories is essential for the development of AI systems and other innovative technologies. This data makes it possible to train algorithms that are able to recognize patterns, make more accurate and earlier diagnoses and optimize treatments, among other things.

The use of medical data therefore promises significant medical progress and the improvement of healthcare for society as a whole.

However, the use of real patient data comes with potential risks. A greater extent to which sensitive medical data is used opens up more opportunities for misuse or unauthorized access. The question of who gets access to the valuable data and how it may be used raises serious ethical and legal questions.

How can we ensure that medical progress is not made at the expense of privacy and data protection? What mechanisms and controls are needed to regulate access to health data so that its use is in line with individual rights and the common good? In other words, how can the data treasure trove be used in a responsible way that respects, among other things, the European data philosophy?

RESEARCH VS. MARKETING

The use of personal health data can be broadly divided into two uses: Focus on medical content for research purposes vs. focus on personal content for marketing purposes. The personal nature of the data plays a decisive role here and has a significant impact on the ethical dimension of this use.
In the field of medical research, the focus is on the medical properties of health data. For example, a brain tumor can be identified on a CT scan, and AI applications can be trained using such health data to assist in the early detection of cancer in radiological imaging. In research, the focus is on general medical progress and the improvement of diagnosis and treatment methods. The personal reference of the data does not play a direct role here, as it is primarily about the medical characteristics and patterns contained in the data.

On the other hand, there is the use of health data for marketing purposes. The focus here is precisely on the personal nature of the data. Systematic user profiling is used to collect specific information about patients' state of health in order to develop purchasing intentions and marketing strategies. Patients' information is analyzed in order to enable companies with a commercial interest to target them as a customer group. The personal nature of the data is of crucial importance here, as it forms the basis for personalized marketing strategies.

The guiding distinction between research and marketing as the intended use of health data plays a crucial role for informed patients. It is understandable that they feel the personal nature of their sensitive data is worth protecting and are weighing up whether they are willing to share their information. The ability to distinguish between pure research use and commercial intent is an important factor that can significantly influence the attitude and willingness to share data and should therefore always be made clear.

However, removing the personal reference in order to conduct research exclusively, without any marketing intentions, is no easy task. The following section goes into more detail and discusses the challenges of ensuring data protection and the anonymity of health data.

PSEUDONYMIZATION VS. ANONYMIZATION

The assumption that complete anonymization of health data was a solution was widespread for a long time. It was assumed that complete anonymity could be guaranteed by carefully removing direct identifiers such as name and date of birth. However, experience has shown that in most cases this is not the case. Even with thorough removal of direct identifiers, the combination of different data points can be used to draw conclusions about a person's identity. Just as a bicycle lock can usually be broken with enough criminal energy, health data can also be re-identified.

This phenomenon is due to the fact that many data points contain information that could allow a person to be uniquely identified. Individual characteristics such as age, gender, zip code or certain medical diagnoses, in combination with other available data sources, can result in the data as a whole representing a globally unique pattern. This process is often referred to as a "linking attack".

For this reason, it is usually necessary to use pseudonymized data. In pseudonymization, personal characteristics are removed or replaced by a code that does not directly reveal the original identity. Pseudonymization is the main starting point of the General Data Protection Regulation (GDPR) for balancing conflicting interests in the context of personal data. Through pseudonymization, affected patients retain their informational self-determination, while at the same time enabling the responsible researchers to process personal health data and opening up scope for action.

Nevertheless, the question of which understanding should be taken as the basis for the personal reference of health data and under which conditions a pseudonymized or anonymized data set can be assumed remains fraught with considerable legal uncertainty. Developments at a technological and legal level are dynamic and require an ongoing examination of these issues in order to ensure adequate protection of privacy and, at the same time, progress in medical research.

ACTIVE VS PASSIVE CONSENT (OPT-IN VS. OPT-OUT)

Regulations on the handling of health data face the challenge of finding an appropriate balance between the protection of patients' informational self-determination on the one hand and the need for data for research and innovation on the other. The so-called opt-in and opt-out procedures play a central role here.

With the opt-in procedure, patients must actively give their consent to the use of their health data for certain purposes. This means that they must explicitly agree to their data being used. This procedure places greater emphasis on the preservation of patients' informational self-determination. The opt-in procedure gives them control over who has access to their data and for what purposes it may be used. Such a consent procedure fully respects privacy and gives patients the opportunity to share their data selectively.

This contrasts with the opt-out procedure, in which the use of health data is automatically accepted unless patients explicitly object. In this case, the data is used by default unless the patient actively objects. The opt-out procedure makes it easier for researchers and innovators to access health data, as they can start from a broader database. However, it also carries the risk that patients give their consent unconsciously or are unaware of how their data is being used. This could lead to a loss of informational self-determination.

In view of the real risk of re-identification and the potential for informational exploitation, the question arises as to whether an opt-out procedure for pseudonymized health data would be fundamentally compatible with the current General Data Protection Regulation (GDPR). The GDPR attaches great importance to the protection of personal data and in most cases requires the active consent of the data subject. In this context, an opt-out procedure could call into question the requirements of the GDPR with regard to the protection of privacy and informational self-determination. A critical examination of the legal and ethical implications is therefore essential in order to find the right way to handle health data.

DATA PROVISION VS. DATA USE

The use of health data for research and innovation offers enormous opportunities and possibilities for medical progress. This form of use is based on a complex value chain that begins with data provision and ends with data use.

Data provision is the fundamental prerequisite for data use. It is crucial that patients can provide their data voluntarily and in an informed manner.

In a highly sensitive area such as healthcare, a simple binary decision between releasing and not releasing data is not enough. Instead, a graduated, differentiated consent procedure is required that meets the individual needs and preferences of patients. The possibility of fine-grained consent plays a decisive role here, allowing patients to specifically determine the purposes for which their data may be used.

Today, modern technologies such as smartphones make it possible to carry out the consent process for data release quickly and well-informed. Patients can retain control over their data via their mobile devices and give their consent specifically for certain research projects or innovations. This strengthens patients' informational self-determination and enables them to handle their health data with confidence.

In addition to consent to data release, the quality of the data is also of crucial importance. Only data of sufficient quality can be used in research and innovation. This requires uniform standards and formats to ensure that the data provided is comparable and interoperable. Only then can it be used effectively and efficiently to gain meaningful insights and develop innovative solutions.
The provision of data is therefore a basic prerequisite for the use of data to promote research and innovation. It requires a differentiated and patient-oriented approach to patient consent as well as uniform standards and formats to ensure high data quality. By creating these conditions, the full potential of health data use can be exploited and medical progress driven forward.

GOVERNMENT IT PROJECT VS. MARKET-BASED SOLUTIONS

The use of health data for research and innovation requires broad cooperation between various stakeholders and the creation of suitable framework conditions. The state plays an important role in creating these basic conditions. For example, in defining standardized and interoperable data formats.

The General Data Protection Regulation (GDPR), which came into force in 2018, also lays down clear rules for the protection of personal data as a basis for data use and focuses on the informational self-determination of patients. The GDPR provides patients with the legal tools to decide what information about them may be made available to whom and under what circumstances. It thus creates a framework that makes it possible to share personal health data in accordance with the interests of the data subjects.

This also forms the basis for private companies to promote the use of health data for research and innovation within the framework of market-based solutions, while upholding the principles of an enlightened and democratic society. On the one hand, there is a risk that the state will restrict patients' autonomy and right to informational self-determination with a paternalistic attitude. A state-imposed opt-out procedure for the secondary use of health data is classified as contrary to the GDPR by leading legal scholars in this field.

The state as an actor in the implementation of IT projects is often subject to difficult balancing of interests, as these are often projects with a political dimension that take into account different party constellations and voter interests. However, these considerations are irrelevant arguments that have a potentially counterproductive effect on the success criteria for the development of a functioning IT infrastructure. Decision-making processes and implementation times for government projects often take years, while the technology has already evolved and changed.

Against this backdrop, the question arises as to whether state IT projects for the provision of healthcare data are in line with the GDPR and the requirements of fast-moving technological developments. Market-based solutions from private companies, which operate in compliance with data protection regulations, could be a sensible alternative in order to promote efficient and modern decentralized solutions that enable medical progress through the use of health data.

SUMMARY

The challenges of using health data for research and innovation are dependent on a social discourse. To this end, two important values-patient informational self-determination and medical progress-are discussed, which seem to collide with each other. The article emphasizes that real patient data is essential for developing artificial intelligence in medicine and driving medical progress. At the same time, however, there is a risk that the use of such data could jeopardize patients' informational self-determination.

Various aspects of this problem are examined. First, the importance of the personal reference for the intended use of health data is discussed. It is emphasized that complete anonymization of health data is not possible in most cases and that pseudonymized data must therefore be used. It is emphasized that differentiated consent and fine-grained control over the use of data are necessary in order to enable patients to make a sovereign decision.

The General Data Protection Regulation (GDPR) is cited as the basic framework for the protection of personal data, which enables private companies to advance the use of health data within the framework of the legal requirements.
Overall, it is clear that the use of health data for research and innovation is a complex task that requires differentiated consideration and balancing of interests.

SOURCES

- “Datenkörper Und Volksgesundheit: Debatte Um Gesundheitsdaten & Datenschutz.” *Kuketz IT-Security Blog*, 2023, www.kuketz-blog.de/datenkoerper-und-volksgesundheit-debatte-um-gesundheitsdaten-datenschutz/.
- “Datenschutz-Grundverordnung (DSGVO).” Dejure.Org, dejure.org/gesetze/DSGVO. Accessed 5 Feb. 2024.
- Gesundheitsdatennutzung – Sicher Und Souverän.” *Acatech*, 29 Jan. 2024, www.acatech.de/publikation/gesundheitsdatennutzung-sicher-und-souveraen/.
- Mühlenbeck, Robin L. *Anonyme Und Pseudonyme Daten.* Nomos, 2023.
- Specht, Louisa. Konsequenzen *Der Ökonomisierung Informationeller Selbstbestimmung: Die Zivilrechtliche Erfassung Des Datenhandels.* Carl Heymanns Verlag, 2012.

ABOUT THE AUTHORS

Sassan Sangsari, MD is a trained physician and philosopher, as well as a Med-Tech entrepreneur. Through his work for ARIC, he deepens his understanding of Artificial Intelligence in medicine.

sangsari@aric-hamburg.de

Dr. Natalie Rotermund is a Ph.D. neuroscientist with many years of experience in the field of science. As a scientific officer at ARIC, she oversees the topics of quantum technologies and AI in medicine and life sciences.

rotermund@aric-hamburg.de

aric