MEDICINE, LAW & SOCIETY
Vol. 17, No. 1, pp. 131–152, April 2024

LEGAL DIFFICULTIES ASSOCIATED WITH THE USE OF BIG DATA IN HEALTHCARE: CIVIL LAW AND CYBERLAW REVIEW

Accepted
17. 12. 2023

Revised
17. 2. 2024

Published
18. 4. 2024

DIYORA IMAMALIEVA, ELNORA INAMDJANOVA, MUKHAMMAD ALI TURDIALIEV, AKMALJON AKRAMOV
Tashkent State Law University, Tashkent, Uzbekistan
diyoraimamalieva@gmail.com, elnorainamjanova100898@gmail.com, m.turdialiyev@gmail.com, a.akramov@tsul.uz

CORRESPONDING AUTHOR
diyoraimamalieva@gmail.com

Abstract Current trends suggest a future in which big data can be used to predict individual health risks and outcomes with exceptional accuracy. By analyzing large datasets that include various health indicators, algorithms can predict the onset of disease, allowing for early intervention and personalized treatment plans. However, the legal and ethical implications of predictive health analytics, such as potential misuse of predictive information or unfair discrimination based on health risks, require careful consideration. In terms of global health surveillance, the COVID-19 pandemic has highlighted the potential of big data to track the spread of disease and inform public health responses.

Keywords
COVID-19,
data ownership,
GDPR,
medical data,
privacy law

https://doi.org/10.18690/mls.17.1.131-152.2024

University of Maribor Press

1 Introduction

Big data has revolutionized many aspects of modern healthcare and continues to be the driving force behind innovative breakthroughs. As health information systems increasingly generate huge pools of data, researchers, clinicians, and policymakers have discovered many opportunities to leverage this resource in a variety of health care applications. For example, big data has found widespread use in predictive modeling, leading to improved diagnoses and proactive health management (Wang, Kung, & Byrd, 2018, p. 4). Using large data sets, predictive models can identify patterns that may not be apparent through human analysis alone, allowing health risks to be identified early. There is active debate about the legality and ethical implications of such predictive models, especially in terms of privacy and consent.

2 Methodology

A methodological approach of this research is based on existing literature on the intersection of Big Data and healthcare law. This involved accessing legal journals, academic publications, government reports, and industry analyses. The review focused on identifying legal challenges, regulatory frameworks, and emerging trends related to the use of Big Data in healthcare settings. We developed a structured framework for analyzing the legal complexities associated with Big Data in healthcare, drawing on principles of civil law and cyberlaw. This framework served as a guide for examining the various legal issues, including data privacy, security, liability, intellectual property rights, and regulatory compliance. Additionally, we analyzed relevant case studies and legal precedents involving the use of Big Data in healthcare to gain insights into how courts have interpreted and applied existing legal principles in this context. This jurisprudential analysis helped to identify key legal doctrines and principles that are relevant to addressing legal challenges associated with Big Data in healthcare.

We conducted a comparative analysis to compare the legal frameworks governing Big Data in healthcare across different jurisdictions. This comparative approach allowed us to identify variations in legal requirements and regulatory approaches, as well as potential best practices for addressing legal difficulties associated with the use of Big Data in healthcare. We synthesized the findings from the literature review, legal analysis, case studies, and comparative analysis to develop a comprehensive understanding of the legal challenges and opportunities associated with the use of Big Data in healthcare. Based on this synthesis, we have formulated recommendations for policymakers, healthcare organizations, legal practitioners, and other stakeholders to effectively navigate the legal complexities of Big Data in healthcare.

3 Legal considerations associated with processing health-related data

The legal status of health-related data is a critical component of the broader legal debate about big data in healthcare. Processing such data requires a thorough understanding of its legal status, especially regarding consent, anonymization, data ownership and data transfer. Next, we will substantively analyze the content of the legal status data through these features.

– Agreement

Health-related data typically falls under the category of sensitive personal data, the collection and use of which requires explicit consent[1]. However, the proliferation of big data complicates the consent process due to the sheer scale and diversity of the data processed, challenging conventional understandings of informed consent (Mittelstadt & Floridi, 2016, p. 7). The question of whether existing consent models are sufficient or whether new models tailored to big data should be developed is an active area of legal debate.

– Anonymization

Anonymization techniques play a key role in protecting privacy when using big data in healthcare. The GDPR, for example, promotes the pseudonymization and anonymization of personal data. However, concerns remain about the potential re-identification of individuals from supposedly anonymous data, a problem that is exacerbated in the context of big data due to the increased likelihood of unanticipated data connections (Rocher, Hendrickx & de Montjoye, 2019, p. 6).

[1] 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) [2016] OJ L 119/1.

– Data Ownership

The issue of data ownership remains legally controversial. The lack of a universal legal framework that clearly defines data ownership leaves the issue subject to varying national laws, contractual agreements, and ethical considerations (Rumbold, & Pierscionek, 2017, p. 2). In the healthcare context, the legal determination of data ownership (whether that of the data subject, a healthcare provider, or a third party) significantly impacts data management and the allocation of rights and responsibilities regarding data use.

– Data transfer

Data transfers, especially international ones, are another important legal consideration. Cross-border transfers of personal data are strictly regulated by international laws such as the GDPR, primarily due to privacy concerns. In the healthcare context, these rules have significant implications for global health research collaborations, multinational healthcare providers, and digital health companies operating in multiple jurisdictions.

The complex legal status of health-related data highlights the need for a comprehensive legal understanding and a robust regulatory framework to ensure privacy protection and ethical handling of data while realizing the enormous potential of big data in healthcare. In the following subsections, we delve deeper into the legal and ethical issues associated with big data, exploring potential regulatory solutions.

4 Complexities of obtaining and de-identifying health-related data

Another distinctive feature of the legal challenges in data collection is the process used to obtain and de-identify health-related data. Collecting health data for big data analytics poses unique legal challenges. The complexities of data collection and anonymization processes have raised significant legal and ethical considerations that require careful attention to data protection laws and regulations. Legal challenges to data collection arise from the need to balance the potential benefits of big data with privacy rights. The collection of health data must comply with laws such as the EU GDPR (2016), which provides for explicit informed consent for the processing of personal health data. However, since big data typically involves a massive and diverse array of data, the process of obtaining informed consent is complex. Thus, the practical implementation of consent requirements in the context of big data is an ongoing area of legal research and debate (Solove, 2013, p. 15).

While anonymization techniques aim to protect privacy by separating data from identifying characteristics, the vastness and diversity of big data pose unique challenges to achieving this goal. Despite the use of advanced anonymization techniques, it may be possible to re-identify individuals by cross-referencing different data sources (Zarsky, 2019, p. 167). Controversy surrounding the effectiveness of anonymization techniques highlights the need for best practices and legal regulations regarding the ability to re-identify based on supposedly anonymized data. Legislation such as GDPR aims to mitigate these problems by introducing strict requirements for data collection, anonymization and other aspects of data processing. However, legislation alone may not be enough, as the rapid development of big data technologies often outpaces the evolution of the legal framework. Ongoing discussions about adapting existing legal frameworks or developing new ones to consider big data are essential to addressing these challenges.

5 The ownership, control and transfer of digital health data

Another issue of data ownership and control is at the center of the legal debate surrounding big data in healthcare. With the increasing volume of health-related data generated from various sources such as electronic health records, wearable devices, and genome sequencing, determining who owns, controls, and can profit from this data remains controversial. In traditional healthcare settings, patient data was often viewed as the property of the healthcare provider or institution that collected it. However, with the proliferation of digital health data, this concept has become increasingly complex (Rumbold & Pierscionek, 2017, p. 7). Some argue that data should be viewed as a natural extension of the person it represents, thereby giving individuals ownership of their health data. However, others argue that health care providers or data processing organizations should retain ownership given their role in collecting, storing, and analyzing data. Different national and international laws offer varying views on this issue. For example, the GDPR does not explicitly mention data ownership, but emphasizes the rights of data subjects, including the right to access, correct, and delete personal data[2]. Meanwhile, in the United States, the Health Insurance Portability and Accountability Act (HIPAA), does not provide patients with data ownership, but does provide them with broad rights to access their health data.[3]

Control of health-related data concerns who has the power to dictate its use. Data control is often related to data ownership, but they are not synonymous. For example, a patient may own their health data but not have full control over it, depending on an agreement with the health care provider or data processing organization. The data control debate largely revolves around the rights to share, sell and repurpose data. Here, ethical considerations often intersect with legal issues, especially regarding informed consent and confidentiality. Typically, patients' explicit consent is required before their data can be used for research or commercial purposes, but the sheer scale and variety of big data makes this process difficult.

One of the major legal issues surrounding big data in healthcare is privacy and data protection. As the amount of health-related data collected and processed increases, ensuring the confidentiality and integrity of this information becomes paramount. Data privacy refers to the right of individuals to control or influence what information is collected about them and how it is used. In the healthcare sector, data privacy is a fundamental aspect of patients' rights and is closely related to the concept of medical confidentiality. In many countries, healthcare providers are required by law to respect and protect patient privacy (Wachter, Mittelstadt & Floridi, 2017, p. 13).

Several international and national laws and guidelines emphasize the importance of data privacy. The GDPR in the EU, for example, enshrines data privacy as a fundamental right and sets strict rules for the processing of personal data. Similarly, in the United States, HIPAA contains provisions protecting patient privacy, particularly with respect to electronic health data. However, the emergence of big data poses new challenges regarding data privacy. Traditional privacy-preserving mechanisms, such as data anonymization, may be inadequate in the face of advanced

[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) [2016] OJ L 119/1.

[3] H. Rept. 104-736 – Health Insurance Portability and Accountability Act of 1996. 104the Congress Report, 2d Session.

data analytics that have the potential to re-identify anonymous data. There is ongoing debate about how existing privacy laws should evolve to address the unique challenges posed by big data.

Data security refers to the measures, policies and tools used to protect data from unauthorized access, data leakage and other forms of data misuse. In the healthcare context, data protection is critical not only to ensure patient privacy, but also to maintain trust and integrity of healthcare systems.

Many countries have specific laws and regulations that define the duties and responsibilities of data controllers (the bodies that determine the purposes and means of processing personal data) and data processors (the organizations that process data on behalf of data controllers) in protecting personal data. Also, both GDPR and HIPAA contain clear data protection requirements, including the need for adequate security measures and breach notification procedures.[4] Despite these legal requirements, data breaches in healthcare remain a serious problem. Factors such as the high cost of medical data on the black market, the rapid digitization of medical records, and the increasing sophistication of cyber threats contribute to this ongoing problem.

The international private law aspect of big data lies in the procedure for data transfer, in particular, navigating the legal landscape of data transfer at the cross-border level deserves special attention. In the era of globalization and digital connectivity, health data often crosses borders, whether for research collaboration, outsourcing of services, or use of cloud storage and processing platforms. International transfers of health-related data add to the legal complexities associated with big data due to differences in national data protection laws and jurisdictional issues.

Regulatory approaches to international data transfers are mostly related to jurisdictional restrictions. Most countries have restrictions on the transfer of personal data outside their jurisdiction. For example, in the European Union (EU), the GDPR stipulates that data transfers to third countries or international

[4] 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) [2016] OJ L 119/1.
H. Rept. 104-736 – Health Insurance Portability and Accountability Act of 1996. 104th Congress Report, 2d Session.

organizations can only occur if the receiving party provides an "adequate" level of data protection. Similarly, in the United States, HIPAA allows covered entities to share protected health information with business associates located abroad, but those associates must comply with the applicable provisions of the HIPAA. Other countries, such as China and Russia, have strict data localization requirements that require personal data to be stored and processed on servers located on their territory[5]. A major challenge in cross-border data transfers is the harmonization of different data protection laws and standards in different countries. This discrepancy can lead to legal uncertainty and hamper global collaboration on health research and services. Several mechanisms have been developed to ensure secure and lawful international data transfer. One of these is the use of Standard Contractual Clauses (SCCs) in the EU, which are contractual terms obliging both the sender and recipient of data to protect personal data.[6]

Another approach is to develop cross-border privacy rules, as seen in the Asia-Pacific Economic Cooperation (APEC) Privacy Framework. This framework encourages the 21 member countries to create a voluntary accountability system for organizations sharing data across national borders.[7]

International transfer of health-related data is an important but challenging aspect of big data in healthcare. Balancing the need for global data flows with the requirement for strong data protection requires a detailed understanding of the associated legal implications.

6 Civil law perspectives on healthcare related big data

Civil law perspectives on big data are examined from the perspective of contractual principles of regulation, ownership of big data, tort relations, and issues of civil liability.

[5] Cybersecurity Law of the People's Republic of China (Effective June 1, 2017), Rogier Creemers, Graham Webster, Paul Triolo, Retrieved from https://digichina.stanford.edu/work/translation-cybersecurity-law-of-the-peoples-republic-of-china-effective-june-1-2017/ (October 9, 2023).

[6] EU Standard Contractual Clauses, Retrieved from https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/standard-contractual-clauses-scc_en (October 9, 2023).

[7] APEC Privacy Framework, (2017) CTI Sub-Fora & Industry Dialogues Groups, Digital Economy Steering Group (DESG).

1) Contract law plays an important role in regulating relationships in the big data environment, especially between healthcare providers, patients, data processors and third parties. Here we explore its relevance and application, focusing on consent, data ownership, responsibility and privacy. One of the core elements of the treaty is the concept of "free and informed consent". In the context of health-related big data, this becomes challenging due to the sheer volume and sensitivity of the data that is collected and processed. Ideally, consent should cover the collection, analysis, storage, sharing and potential future use of data. However, fully informed and specific consent becomes difficult when the future use of the data is unclear at the time of collection. This dilemma has sparked debate about the concept of "broad consent," where people agree to unspecified future uses of their data.

Contract law is also critical to defining the rights and responsibilities of data ownership, which can be ambiguous in the big data space. These contracts typically specify who owns the data, who has the right to use and access it, and under what conditions. These definitions are critical because they have implications for intellectual property rights, commercial use, and liability issues (Rush et. al., 2022). In any contractual relationship, determining liability for breach of contract or damage is critical. Given the risks associated with data breaches and misuse, contracts should clearly outline liability and risk management mechanisms. These include provisions regarding data protection, security measures, indemnification provisions and insurance claims (Van Dijck & Poell, 2016, p. 7). Confidentiality clauses in contracts protect the privacy of individuals whose data is collected, stored and analyzed. In the context of big data, these points should cover scenarios such as access to third party data and potential breaches. Ultimately, contract law provides a vital framework for establishing the rights and obligations of parties related to big data. However, due to the new and complex nature of big data, existing contractual models must evolve to adequately address these unique challenges.

2) The primary function of tort law is to provide remedies for types of damage or injury that are not covered by contract law. In the context of big data, it specifically covers cases of negligence and standards of care. The tort of negligence can have various consequences in the field of big data. For example, if a healthcare provider fails to properly protect their patients' data, resulting in a security breach, they may be held liable for negligence. Essentially, this requires establishing a "duty of care" between the data processor and the person whose data has been compromised.

However, the challenge lies in defining the "standard of care" in the context of big data. Given the modernity of the field, it may be unclear what constitutes "reasonable" data practices. Therefore, the definition of this standard requires serious consideration and discussion in a legal, ethical and technological context. Another pressing issue in the application of data in tort regulation is damage caused by misuse or mishandling of data. One of the essential elements of a negligence claim is that the plaintiff must have suffered "damage." In data breach cases, demonstrating harm can be challenging. While some harms, such as identity theft or financial loss, are tangible, others, such as psychological distress or potential future harm, may be more difficult to prove (Solove & Citron, 2016, p. 746). However, emerging case law and legislative changes in various jurisdictions are beginning to recognize these intangible harms. Courts increasingly agree that people suffer real harm when their personal information is exposed, even if the harm is not yet realized (Deucher, 2023, p. 50).

In a big data ecosystem, data often passes through multiple hands—health care providers, data analysts, third-party vendors, etc. This raises issues of vicarious liability—where one party is liable for the negligence of another (Miller & Weckert, 2018, p. 9). Who should be held responsible if a third-party vendor experiences a data breach? Such questions are becoming increasingly relevant and complex in the field of big data. As demonstrated, tort law offers a valuable framework for understanding and addressing harm and negligence in the context of big data. However, the unique challenges and complexities associated with big data require constant refinement and adaptation of these traditional legal principles.

3) Big data in healthcare, despite its enormous potential, also poses significant risks, especially in terms of civil liability. This subsection addresses civil liability issues in the collection, storage, and use of big data in healthcare. The first step in any big data project is data collection. In the healthcare industry, this data is often sensitive and personal. Civil liability may arise if data is collected without the necessary consent or if the data collection violates privacy laws. For example, large fines can be imposed for illegal data collection under the EUs (GDPR).[8]

[8] 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) [2016] OJ L 119/1.

Civil liability can also arise from how medical data is stored. If data is stored in a way that makes it vulnerable to data breaches, and such a breach occurs, the person responsible for storing the data may be liable for any damages caused to the individuals whose data was breached. This has been seen in various data breach cases where companies have faced lawsuits due to unauthorized access to their customers' data (Riley, 2023). Finally, liability may arise from the use of collected and stored data. For example, using patient data for purposes other than those for which consent was given could potentially result in civil liability. In addition, decisions made based on big data analytics may result in liability if they result in harm. To give an example, if an algorithm used in healthcare makes a recommendation that results in inappropriate treatment and subsequent harm to a patient, there may be grounds for a liability claim (Price & Cohen, 2019, p. 8).

To address these civil liability issues, healthcare institutions and other stakeholders in the big data space need to ensure strict compliance with relevant laws and best practices. This may include obtaining clear and informed consent for data collection, ensuring strong data security measures, and using data responsibly and ethically. While big data in healthcare offers many promising opportunities, it also comes with significant civil liability risks. Navigating them requires a deep understanding of the legal landscape and ethical data practices.

7 The application of property rights to health-related data

Big data in healthcare can be viewed as an asset, and this raises the issue of ownership.

Personal data as property?

One of the questions that courts have grappled with is whether personal data can be considered property. Some legal systems have approached this idea through the lens of property rights. In the UK, a case known as *Your Response Ltd v Datateam Business Media Ltd* (2014)[9] held that databases could be considered property, but did not deal with the data in the database as such.[10] However, most legal systems do not provide

[9] Case *Your Response Ltd v Datateam Business Media Ltd* [2014] EWCA Civ 281 (14 March 2014).

[10] Case *Your Response Ltd v Datateam Business Media Ltd* [2014] EWCA Civ 281, [2015] Q.B. 41

full ownership rights to personal data due to concerns about its commodification and the implications for individual rights and dignity (Custers & Malgieri, 2022, p. 3).

Medical data: a special case?

Medical data may be a special case. In the US, *University of Washington v. Catalona* (2007) held that individuals do not retain ownership of their biological materials once they have been voluntarily provided for research (*University of Washington v. Catalona* [2007] WL 1758268). This case has implications for big data in healthcare because it suggests that data derived from such materials can also be considered outside of human property rights. However, this solution is not generally accepted and is the subject of criticism and debate (Rodwin, 2010, p. 12).

On the other hand, healthcare providers and organizations that collect and store patient data typically do not have ownership rights to that data. In the United States, HIPAA gives patients significant rights to access their data, presuming that it is not "owned" by the provider.[11] However, providers may have certain rights in the systems and databases they use to store and organize this data. In many cases, ownership rights associated with Big Health Data are determined by contracts. When patients provide data for research, they typically sign agreements that outline the conditions under which the data will be used. These agreements may determine who has the right to use the data, for what purposes, and under what conditions. However, the enforceability of these contracts may be subject to legal review, especially if they are considered unfair or if the patient has not given informed consent (Andreotta, Kirkham & Rizzi, 2022, p. 10). The law in this area continues to evolve, and the outcome of future court cases is likely to have a significant impact on how these issues are resolved.

8 Case Studies: Exploring Civil Cases Related to Big Data

The use of big data in healthcare has resulted in several landmark cases that highlight the complexities and challenges of navigating this new territory. This section provides an overview of some important civil cases involving big data.

[11] H. Rept. 104-736 – Health Insurance Portability and accountability act of 1996. 104th Congress report, 2d Session.

Example 1: *Google DeepMind and the Royal Free London NHS Foundation Trust*

In 2017, the UK Information Commissioner's Office (ICO) ruled that the Royal Free London NHS Foundation Trust did not comply with the Data Protection Act when it provided patient data to Google DeepMind. The Trust provided the personal data of around 1.6 million patients as part of a trial of the Acute Kidney Injury Alert, Diagnosis and Detection System. The ICO found that patients were not adequately informed about the use of their data, which constituted a serious breach of patient privacy. The shortcomings found by the ICO breached the following data protection principles:[12]

Principle One: Personal data shall be processed fairly and lawfully;

Principle Three: Personal data should be adequate, relevant and not excessive;

Principle Six: Personal data shall be processed in accordance with the rights of data subjects;

Principle Seven: Appropriate technical and organizational controls shall be taken – this includes the need to ensure that appropriate contractual controls are in place when a data processor is used.

Example 2: *IMS Health Inc. v. Sorrell*

In the United States, the Supreme Court case IMS Health Inc. v. Sorrell (2011)[13] played a key role in defining the limits of data privacy in healthcare. The court ruled that a Vermont law restricting the sale, disclosure, and use of pharmacy records violated the First Amendment (*Sorrell v. IMS Health, Inc.,* [2011] 564 U.S. 552). This case highlighted the tension between personal privacy and commercial interests in the use of medical data.

[12] Royal Free and Google DeepMind trial did not comply with DPA, Retrieved from https://www.digitalhealth.net/2017/07/royal-free-and-deepmind-did-not-comply-with-dpa-ico/ (October 9, 2023).

[13] Case *Sorrell v. IMS Health, Inc.*, 564 U.S. 552 (2011), retrieved from https://supreme.justia.com/cases/federal/us/564/552/.

These cases highlight the legal complexities of using big data in healthcare, with issues surrounding patient consent, data privacy, anonymity and commercialization coming to the fore. As we move into the uncharted waters of big data in healthcare, these issues will continue to challenge legislators, healthcare providers and patients.

9 Data protection rules: application to big data

In the context of cyber law, data protection rules are of paramount importance for the use of big data, especially in the healthcare sector where personal health information is involved. As a result, countries have taken different positions regarding data protection. This subsection will review and discuss current data protection regulations in the G7 countries - Canada, France, Germany, Italy, Japan, UK and US - and their application to big data in healthcare.

Canada*:* Canada's Personal Information Protection and Electronic Documents Act (PIPEDA) regulates how private sector organizations collect, use and disclose personal information in the course of business. This also applies to health information, and any use of big data in healthcare must comply with the PIPEDA principles, particularly regarding consent and reasonable purpose. PIPEDA’s main goal is to regulate private-sector organizations across Canada that collect, use or disclose personal information in the course of a commercial activity. The law defines a commercial activity as any particular transaction, act, or conduct, or any regular course of conduct that is of a commercial character, including the selling, bartering or leasing of donor, membership or other fundraising lists.[14]

France*:* In France, the Commission Nationale de l' Informatique et des Libertes (CNIL) acts as a supervisory authority in the field of personal data protection, and monitors the law to ensure it is implemented correctly (advices and recommendations). It also provides opinions on the legality of data processing (authorization requests), participates in jurisdictional appeals in case of violation of the law, and controls the entire process.[15]

[14] The Personal Information Protection and Electronic Documents Act (PIPEDA), Retrieved from https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/ (October 9, 2023).

[15] The Law "Informatique et Libertés", French Act No. 2018-493 of 20 June 2018. Retrieved from

In addition, the CNIL has discretion both to control the process of initiating processing and to impose sanctions (for example, warnings to stop processing data and imposing financial penalties). In its activities, CNIL applies the EU GDPR to big data. Thus, big data in healthcare must comply with the principles of the GDPR, including the lawful basis for processing, data minimization and data subject rights.

Germany*:* Germany also applies the GDPR, but has specific national legislation for health data under the Federal Data Protection Act (BDSG). However, with regard to the processing of medical data, Germany adopted a special law on the protection of patient data, Patientendaten-Schutz-Gesetz (PDSG), dated from 2020. The PDSG applies to all healthcare providers, including hospitals, doctors, health insurance providers and pharmacies, that use services, applications and telematics infrastructure components of the German healthcare system to process patient information.[16] According to the new law, starting from 2021, health insurance providers are required to offer electronic patient files (records) or Elektronische Patientenakte (ePA) to customers. From 2022, ePAs will also include sensitive information that has until now only been documented in hard copy, such as pregnancy and birth records, children's medical records and vaccination records. Patients will have the right to decide what is stored in their EPA and who has access to it.

Starting in 2023, patients will be able to voluntarily provide their EPA data to researchers through a “data donation”—i.e. as part of the free voluntary transfer of personal data. Patients will have to give their informed consent, but they will be able to do so digitally. Data donations are limited to specific research purposes, such as improving the quality of healthcare. Patients will also be able to choose how much of their data is donated and limit access to certain information.[17]

https://caseguard.com/articles/the-new-data-privacy-and-protection-landscape-in-france/#:~:text=2018%2D493%20of%2020%20June%202018%20implements%20the%20provisions%20of,collected%20and%20processed%20within%20France (October 9, 2023).

[16] Draft bill from the federal government Draft of a law to protect electronic patient data in the telematics infrastructure (Patientendaten-Schutz-Gesetz-PDSG). BT-Drs. 19/18793.

[17] Working Group Report on Virtual Health and Care. The Future of Virtual Health and Care: Driving access and equity through inclusive policies, 2022, p. 104.

Japan: Japan's Law on the Protection of Personal Information (APPI) applies to big data in health care, requiring the lawful and fair acquisition of personal data and taking into account the rights of the data subject. The Japanese APPI sets strict requirements for data processors, such as the need for individual consent in the case of sensitive information.[18] Since medical data often falls into this category, big data operations in Japan must carefully adhere to these regulations.

UK*:* Although the UK has left the EU, it continues to adhere to the principles of the GDPR under the Data Protection Act 2018. The Information Commissioner's Office provides guidance on the use of big data, artificial intelligence and machine learning.[19] The UK GDPR guidance is intended for those with day-to-day data protection responsibilities. It explains the general data protection regime that applies to most UK businesses and organizations. It covers the UK GDPR, developed under the Data Protection Act 2018. It explains all principles, rights and obligations of data protection. This guidance also contains links to more detailed guidance and other resources, including ICO guidance and ICO statutory codes of practice. Links to relevant guidance published by the European Data Protection Board (EDPB) are also included for reference purposes.[20]

US*:* The US does not have a federal data protection law similar to the GDPR. Instead, it has industry-specific privacy laws such as HIPAA for health information. HIPAA privacy and security regulations play a central role in the use of big data in healthcare.[21] It applies to the collection of information in hospitals, doctors' offices and other places where health care services are provided, as well as in business enterprises that help service providers manage and store data.

HIPAA is based on two important ideas in patient care: privacy and confidentiality. Privacy refers to a person's right to limit who knows what about their medical condition. This also includes the right to have conversations about health care conducted in places where they cannot be overheard by others. The detailed regulation that HIPAA covers is called the Privacy Rule. Confidentiality refers to the

[18] Act on the Protection of Personal Information Act No. 57 of (2003).

[19] Guide to the General Data Protection Regulation (GDPR), Information Commissioner's Office, 14 October 2022 - 1.1.17, p. 357.

[20] Information Commissioner's Annual Report and Financial Statements 2022/23, July 2023 HC 1440.

[21] H. Rept. 104-736 – Health Insurance Portability and Accountability act of 1996. 104th Congress Report, 2d Session.

healthcare professional's duty to maintain information so that it is not disclosed without the patient's consent, unless required by law or considered necessary for clinical reasons.

On the whole, each G7 country has a complex and unique legal environment for data protection, which directly impacts the use of big data in healthcare. Understanding these differences is critical for international health data initiatives and global digital health companies.

10 Privacy laws and big data: impact on data collection and use

In the big data and healthcare industries, privacy laws have a major impact on how information is collected and used. These laws vary widely across jurisdictions, affecting healthcare providers, patients, and big data analysts. In Europe, the GDPR is a comprehensive regulation that affects how big data can be collected and used. The GDPR includes provisions such as the right to be forgotten and data portability, which directly impact how big data is managed. In particular, its principle of "privacy by design and by default" implies that systems processing big data must consider privacy at every stage of data processing (Shay, 2023).

In the United States, HIPAA sets the benchmark for protecting confidential patient data. Under the HIPAA Privacy Rule, personal health information can only be used for medical purposes with appropriate security measures in place. This impacts how health data is processed, especially when converted to big data, and requires organizations to take steps to properly de-identify protected health information.

In Canada, PIPEDA requires consent for the collection, use and disclosure of personal information in the course of business activities, which may include some forms of big data analytics. This requires careful planning and transparency on the part of organizations involved in health-related big data.

The Japanese APPI sets strict requirements for data processors, such as the need for individual consent in the case of sensitive information.[22] Since medical data often falls into this category, big data operations in Japan must carefully adhere to these

[22] Act on the Protection of Personal Information (Act No. 57 of 2003).

regulations. Therefore, privacy laws are vital to the discussion of big data in healthcare. They ensure patient privacy and trust, and pose challenges to maximizing the potential of big data in healthcare. Information security is of utmost importance when dealing with big data in healthcare due to the confidentiality and personal nature of health-related data. Various international and national standards set out the obligations of the parties involved in the processing of such data.

SO/IEC 27001: Information security management: this international standard defines the requirements for an organization to establish, implement, maintain and continuously improve an information security management system. It offers a systematic and structured approach to managing information so it remains secure, spanning people, processes and IT systems.

NIST Cybersecurity Framework (CSF)*:* In the United States, the NIST CSF provides a policy framework of computer security recommendations for private sector organizations to assess and improve their ability to prevent, detect, and respond to cyber incidents.[23] It is a risk-based approach to cybersecurity risk management that is widely applicable to healthcare big data organizations.

Information Technology Act 2000 (ITA-2000)*:* In India, the ITA-2000 provides that entities handling sensitive personal data must maintain reasonable security measures and, in the event of a data breach, they must demonstrate that they have implemented such security measures. This law sets the cybersecurity standards that big data processors in India must adhere to.

Cybersecurity Law of the People's Republic of China (CSL):[24] In China, the CSL is the main law governing cybersecurity and data privacy. This law obliges network operators to ensure network security in accordance with state regulations and mandatory standards. It contains strict rules affecting the processing of big data and imposes significant fines for non-compliance (Creemers, Webster, Triolo, 2018). Essentially, these information security standards define a detailed list of obligations for parties working with big data, especially in the healthcare industry. They provide technical and administrative measures to ensure the confidentiality, integrity and

[23] NIST Releases Cybersecurity Framework 2.0 Draft & Implementation Examples, 2023, Retrieved from https://csrc.nist.gov/News/2023/nist-releases-cybersecurity-framework-2-0-draft (October 12, 2023).

[24] Cybersecurity Law of the People's Republic of China (Effective June 1, 2017).

availability of data. It is crucial for organizations managing big data to become familiar with these standards to avoid legal implications and ensure the secure handling of sensitive health data.

The following cases provide real-life examples of legal issues arising from the misuse of big data in the healthcare sector.

1. Anthem Inc. Data Breach (2015). In 2015, Anthem Inc., one of the largest health insurance companies in the United States, suffered a major data breach. Attackers gained unauthorized access to a company's IT system and stole the personal information of nearly 78.8 million people, including names, Social Security numbers, medical IDs, addresses, and employment information. This incident led to numerous lawsuits against Anthem. In 2018, the company agreed to a $115 million settlement—the largest in the history of the data breach—to provide victims with two years of credit monitoring.[25]

2. Google Project Nightingale (2019): In 2019, it was reported that Google had gained access to the health data of millions of Americans through its Project Nightingale, a partnership with Ascension, one of the largest healthcare systems in the US. The project attracted scrutiny because neither patients nor doctors were informed about the data exchange. This has raised serious questions about HIPAA compliance and consent requirements (Baric-Parker & Anderson, 2020, p. 5). The incident prompted an investigation by the Office of Civil Rights of the US Department of Health and Human Services.

3. SingHealth Data Breach (2018). In 2018, SingHealth, Singapore's largest healthcare group, suffered a major cyber-attack.[26] The personal data of about 1.5 million patients was stolen, including Singapore Prime Minister Lee Hsien Loong. This incident resulted in a fine of S$1 million under Singapore's Personal Data Protection Act (PDPA).

[25] *In re Anthem*, Inc. Data Breach Litigation, 162 F. Supp. 3d 953 (N.D. Cal. 2016).

[26] Singapore Health Services Pte. Ltd. & Ors. [2019] SGPDPC 3.

11 Conclusion

These case studies illustrate the urgent need for a strong legal and ethical framework to protect big data in the healthcare sector. They also highlight the potential consequences of inadequate data protection measures, both in terms of financial penalties and loss of public trust. The examined regulatory frameworks revealed the role of big data in healthcare, including its current applications, impact on patient care, and potential future applications. We also presented case studies that illustrate the different ways big data can be used in healthcare.

References

Andreotta, A. J., Kirkham, N. & Rizzi, M. (2022). AI, big data, and the future of consent. *AI & society*, 37(4), 1715–1728.

Baric-Parker, J., & Anderson, E. E. (2020). Patient Data-Sharing for AI: Ethical Challenges, Catholic Solutions. *The Linacre quarterly*, 87(4), 471–481.

Creemers, R., Webster, G., Triolo, P. (2018). Translation: Cybersecurity Law of the People's Republic of China (Effective June 1, 2017), Digichina, retrieved from https://digichina.stanford.edu/work/translation-cybersecurity-law-of-the-peoples-republic-of-china-effective-june-1-2017/ (October 9, 2023).

Custers, B. H. M. & Malgieri, G. (2022). Priceless Data: Why the EU Fundamental Right to Data Protection is at Odds with Trade in Personal Data. *Computer Law & Security Review*, 45, 1-13.

Deucher, C. M. (2023). Data Breach Standing: How Plaintiffs May Find Their Footing After TransUnion v. Ramirez. *OSLJ Online*, 84 (2), 37-53.

Miller, S. & Weckert, J. (2018). Privacy, the workplace and the internet. *Journal of Business Ethics* 28 (3), 255 - 265.

Mittelstadt, B. D., & Floridi, L. (2016). The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and engineering ethics*, 22(2), 303–341.

Price, W. N., 2nd, & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature medicine*, 25(1), 37–43.

Riley T. (2023). IBM, Johnson & Johnson Hit With Second Health Data Breach Suit. Bloomberg Law Subscription, retrieved from https://news.bloomberglaw.com/privacy-and-data-security/ibm-johnson-johnson-hit-with-second-health-data-breach-suit (October 9, 2023).

Rocher, L., Hendrickx, J. M., & de Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature communications*, 10(1), 1-9.

Rodwin, M. (2010). Patient Data: Property, Privacy & the Public Interest. *American journal of law & medicine,* 36, 586-618.

Rumbold, J. M. & Pierscionek, B. (2017). The Effect of the General Data Protection Regulation on Medical Research. *Journal of medical Internet research*, 19(2), 3069. doi: 10.1038/s41467-019-10933-3.

Rush, M. A., Lawrence, J. H., Rybicki, D. C. & Musselman, L. A. (2020). U.S. Investigations, Enforcement, and White Collar Alert, retrieved from https://www.klgates.com/COVID-19-Looming-False-Claims-Act-Liability-for-Paycheck-Protection-Program-Loans-04-09-2020 (October 12, 2023).

Shay, D. (2023). GDPR Top Ten #6: Privacy by Design and by Default, retrieved from https://www2.deloitte.com/ch/en/pages/risk/articles/gdpr-privacy-by-design-and-by-default.html, (October 9, 2023).

Solove, D. (2013). Introduction: Privacy Self-Management and the Consent Dilemma. *Harvard Law Review*, 126(7), 1880-1903.

Solove, D. J.& Citron, D. K. (2016). Risk and Anxiety: A Theory of Data Breach Harms. *Texas Law Review*, 96(4), 737-786.

Van Dijck, J. & Poell, T. (2016). Understanding the promises and premises of online health platforms. *Big Data & Society*, 3(1), https://doi.org/10.1177/2053951716654173.

Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. *Science Robotics*, 2(6), doi: 10.1126/scirobotics.aan6080.

Wang, Y., Lung, L. A. & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3-13.

Zarsky, T. (2019). Privacy and Manipulation in the Digital Age. *Theoretical Inquires in Law*, 20(1), 157-188.

Povzetek v slovenskem jeziku

Trenutni trendi nakazujejo prihodnost, v kateri se bodo velepodatki z izjemno natančnostjo uporabljali za napovedovanje individualnih zdravstvenih tveganj in izidov. Z analizo obsežnih zbirk podatkov, ki vključujejo različne zdravstvene kazalnike, lahko algoritmi napovedujejo začetek bolezni, kar omogoča zgodnje posredovanje in personalizirane načrte zdravljenja. Vendar pa pravne in etične posledice prediktivne analitike zdravja, kot so možna zloraba prediktivnih informacij ali nepravična diskriminacija na podlagi zdravstvenih tveganj, zahtevajo previdno obravnavo. V smislu globalnega nadzora zdravja je pandemija COVID-19 poudarila potencial velepodatkov za sledenje širjenju bolezni in informiranje odzivov javnega zdravja.