The Risks of Health Data Commodification in the EU Digital Market

by

Chiara Gallese*

CONTENTS

* Chiara Gallese, Marie Sklodowska-Curie Postdoctoral Fellow, Department of Law, University of Turin, Italy; Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, Netherlands; School of Engineering, Carlo Cattaneo University – LIUC, Varese, Italy; ISLC fellow, University of Milan, Milan, Italy; e-mail: chiara.gallese@unito.it; ORCID: https://orcid.org/0000-0001-8194-0261.

Suggested Citation: Chiara Gallese, 'The Risks of Health Data Commodification in the EU Digital Market' (2024) 17 YARS pp. 89–126.

Article received: 29 November 2023, accepted: 13 May 2024.

Yearbook of Antitrust and Regulatory Studies, Vol. 17, No. 29, pp. 89–126 • ISSN 1689-9024
https://doi.org/10.7172/1689-9024.YARS.2024.17.29.3

Abstract

This article explores the health data commodification phenomenon in the European Union's digital market. The emergence of a health marketplace, and the increasing utilisation of health data by both public and private entities, have raised significant concerns about citizens' rights. This study examines the risks of health data commodification, the EU's efforts to facilitate data sharing and reuse through the Digital Strategy Corpus of law (DSC) as well as its potential implications on the rights of data subjects. The article investigates the ongoing scholarly debate surrounding the commodification of personal data, and its ethical and legal dimensions. The interdisciplinary approach intertwines legal analysis with computer science insights to explore re-identification risks. The article highlights the need for a balanced approach that upholds citizens' privacy rights while enabling responsible data sharing.

Résumé

Cet article explore le phénomène de marchandisation des données de santé sur le marché numérique de l'Union européenne. L'émergence d'un marché de la santé et l'utilisation croissante des données de santé par des entités publiques et privées suscitent de fortes préoccupations concernant les droits des citoyens. Cette étude examine les risques de marchandisation des données de santé et les efforts de l'UE pour faciliter le partage et la réutilisation des données grâce au corpus juridique de la stratégie digitale (DSC), ainsi que ses implications potentielles sur les droits des personnes concernées. L'article examine le débat scientifique actuel sur la marchandisation des données personnelles et ses dimensions éthiques et juridiques. L'approche interdisciplinaire mêle analyse juridique et savoir dans le domaine de l'informatique pour explorer les risques de réidentification. L'article souligne la nécessité d'adopter une approche équilibrée qui respecte la vie privée des citoyens tout en permettant un partage de données responsable.

Key words: Health Data; Data Protection; Digital Strategy Corpus; Data Commodification; Anonymisation Techniques; EU Regulation; Privacy Rights.

JEL: K24

I. Introduction

The economic valuation of health data has become a feature of contemporary society. In recent years, health data has transcended its traditional role as a tool to deliver healthcare to patients, and evolved into a powerful economic asset exploited by many different entities ranging from private corporations to public institutions. This paradigm shift has launched the creation of what is commonly referred to as the "health data marketplace",[1] where data is traded, accumulated, and monetised with little attention to data protection and patients' opinions.[2]

The implications of this transformation are not confined to businesses and companies; they extend to the very core of our society, affecting notions of privacy, consent, dignity, and the delicate balance between societal progress and individual rights.[3] Even public institutions now leverage health data for an array of applications, such as disease control, pandemic response, public health expenditure estimation, and machine learning model training.

The European Commission has become a player in the health data marketplace through the creation of a common European Health Data Space (hereinafter: EHDS).[4] This initiative is aimed at "empowering individuals through increased digital access to and control of their electronic personal health data, at the national level and EU-wide, and support to their free movement, as well as fostering a genuine single market for electronic health record systems, relevant medical devices and high-risk AI systems (primary use of data); providing a consistent, trustworthy and efficient set-up for the use of health data for research, innovation, policy-making and regulatory activities (secondary use of data)".[5]

However, the vulnerability of citizens in this context is an issue that must be considered. While these developments are very important to society, the

[1] Jane Thomason, 'Big tech, big data and the new world of digital health' (2021) 5(4) Global Health Journal 165.

[2] Carina Dantas and Karolina Mackiewicz, 'Are we ensuring a citizen empowerment approach for health data sharing?', in Anto Čartolovni and others (eds.) *Proceedings of the 2022 GoodBrother International Conference on Privacy-friendly and Trustworthy Technology for Society* (Zagreb 2022) 55.

[3] Jamie Pinchot, Adnan A. Chawdhry and Karen Paullet, 'Data Privacy Issues in The Age of Data Brokerage: an Exploratory Literature Review' (2018) 19(3) Issues in Information Systems 92; Laura DeFrancesco and Ariel Klevecz, 'Your DNA broker' (2019) 37(10) Nat Biotechnol 842-7.

[4] Giorgia Bincoletto, 'The EDPB-EDPS Joint Opinion on the Commission Proposal for a Regulation on the European Health Data Space: Key Issues to Be Considered in the Legislative Process' (2022) 8 Eur. Data Prot. L. Rev. 398.

[5] See European Commission, 'European Health Data Space' <https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en> accessed 30 September 2023.

exploitation of health data represents an actual risk to individuals' fundamental rights, as the new rules are deeply interconnected with the General Data Protection Regulation[6] (hereinafter: GDPR). The GDPR is a fundamental legal instrument to protect EU citizens' fundamental rights thanks to its core principles, such as data minimisation, purpose and storage limitation, and the mandatory presence of a legal basis. Complementing the GDPR is Regulation EU 2018/1725[7] and Convention 108+[8], each contributing to the complex data protection legal framework. The corpus of national health data laws intersects with many EU laws, often generating legal ambiguity and fragmentation across the Member States.

Health data, classified by the GDPR as a special category of personal data, undergoes several transformations before it is shared and exploited. Typically, techniques such as pseudonymisation, de-identification, and anonymisation are employed as security measures, aiming to cancel or hide personally identifiable information. While these techniques mitigate privacy risks, at least to a certain extent, a problem emerges when data is considered anonymised.

This article argues that anonymised datasets – processed with a variety of anonymisation techniques – stand on the edge of a legal and ethical precipice, and are not correctly protected by existing legal framework. Once classified as anonymised, in fact, these datasets cease to be considered personal data; consequently, the legal safeguards granted to personal data are not present anymore. This legal ambiguity renders them susceptible to commercial exploitation, most often without due consideration of the rights and expectations of data subjects. However, research has shown how true anonymity does not exist in reality, and the concept of anonymisation is not black-and-white; technological progress has made it very easy to re-identify data, even after anonymisation techniques have been applied. The very act of anonymisation, designed to protect privacy, can lead to a legal void that exposes data subjects to several risks, as detailed in the following sections.

---

[6] 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.

[7] Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC [2018] OJ L 295/39.

[8] Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108), as amended by the Amending protocol to the Convention for the Protection of Individuals with Regard to the Processing of Personal Data, adopted by the Committee of Ministers of the Council of Europe at its 128th Session in Elsinore on 18 May 2018.

The insufficiency of scholarly attention to these themes calls for an interdisciplinary approach. This article stems from the author's MSCA project DataCom,[9] which aims to create a new EU framework for the ethical reuse of health data through the confluence of legal analysis, ethics, and technical insights from computer science.

This article is structured as follows, the first section introduces the challenges to be analyzed and the DataCom project as a proposed solution. The second section delineates the literature on the topic and the legal framework. The third section explores problems surrounding the views expressed in case law concerning the concepts of personal data, how they conflict with the concept of technical research, and analyzes how this discrepancy ends up leaving a legal gap that endangers citizens' fundamental rights. The fourth section presents specific cases showing the actual risks posed by the commodification of health data. The fifth section sheds light on the novelties introduced by the EHDS and the EU Artificial Intelligence Act (hereinafter: AIA). The sixth section analyzes the inadequacy of the Digital Strategy Corpus of law (hereinafter: DSC) when it comes to protecting citizens from the risks of health data reuse.

1. A proposal for a new EU framework for the ethical commodification of health data

DataCom is a Marie Sklodowska Curie post-doctoral project financed by the European Union. It advances the debate on health data commodification in the public sector by investigating how the new DSC deals with respecting fundamental rights, how it is coordinated with the existing data protection laws, regulations, and ethical principles, and how to mitigate the conflicts and legal gaps involving citizens in the process. Some studies and European Data Protection Board (hereinafter: EDPB) opinionshave, in fact, raised concerns about the potential conflicts between the DSC, the GDPR, and other regulations. However, the extent to which the recent proposals may exacerbate the commodification of health data within the public sector has not been analyzed. Studies focusing on the risks of the exploitation by public actors of citizens' health information, even when anonymised, are still missing.

There is still a legal vacuum around many topics related to this theme, such as the extent to which public bodies and researchers are allowed to re-share health data with other institutions under the Data Act framework, the FAIR principles, and the Open Data Directive, and the consequences for them and for the data subjects.

[9] See European Commission, 'A new EU Framework for an Ethical Re-use of Health Data' <https://cordis.europa.eu/project/id/101108151> accessed 30 September 2023.

DataCom will thus investigate the boundaries between the exploitation by public bodies of health data for the public good, and the limitations provided by the relevant data protection law, ethical principles, and expectations of citizens, in order to find a balance between them.

DataCom will fill these gaps and push such debates further by exploring to what extent the existing legal framework is inadequate to regulate health data sharing, by investigating how it can be improved to meet the actual needs of the public sector and, at the same time, to protect data subjects' rights and freedoms adequately. By studying the impact of the legal gaps left by the recent EU regulatory proposals, DataCom will also suggest possible solutions to EU policy-makers. DataCom will bridge this knowledge with citizens' needs and expectations towards their data.

For the first time, DataCom builds on an interdisciplinary approach, combining insights from computer science and law, which is indispensable for the study of data protection, as it makes it possible to capture its legal dimension as well as its technical implementation.

It will then investigate how the main actors in the public sector operate with respect to health data processing. It is argued that studying, through the lenses of legal scholars, the way in which public bodies exploit health data and perform the anonymisation process, allows for a better understanding of the data commodification phenomenon. It is also argued that, due to lack of training and awareness, many public servants, including researchers, do not usually perform an ethical assessment of the reuse of health data, and on the compatibility of further processing. This increases the risk of adverse effects, including that of data breaches and re-identification due to inappropriate anonymisation.

While assessing the attitude within the public sector, the gender and diversity dimensions will also be investigated, exploring how the reuse of health data may have an impact on women, people with disabilities, and minorities.

This research is important not only for legal researchers and public servants dealing with health data sets, but also for medical practitioners, university support staff, students, lawyers, national Data Protection Authorities (hereinafter: DPA), and EU policy-makers.

The project will bridge the needs of the public sector with the expectations of citizens regarding the reuse of their data, creating the "New EU Framework of Ethical Commodification". It will consist of a set of theories, rules, and best practices to be employed by public institutions (including hospitals) when re-using health data sets, even if anonymised. It will take into consideration the needs, expectations, and wishes of data subjects (regardless of the fact whether they donated their data as part of the Data Altruism mechanism or not) and perform an ethical assessment before the reuse.

The project will identify shortcomings and gaps with regard to public servants' attitudes towards anonymisation and citizens' needs and will combine those results in order to assess in what area, and to what extent, the existing regulatory framework is insufficient to promote responsible reuse and exploitation of data.

However, the project only focuses on the public sector, while the private sector still needs attention. This new ethical framework might then represent a starting point to build a novel way to regulate data sharing in the private sector, particularly the vast health data market. As data intermediaries play a major role in this complex international landscape, a new EU regulation targeting the sale of health data and fair practices might be the next horizon of the DSC.

Emerging within the contemporary scholarly debate is an initiative that promotes the increase of citizen agency over health data, through the empowerment of individuals to determine what data they choose to monetise. This trend is increasingly gaining popularity, which represents a paradigm shift towards more participatory data governance. However, even as these initiatives evolve, various ethical concerns remain, particularly when considering the following factors.

First, the problem of obtaining genuine and informed consent is a concerning issue. Given the vulnerable nature of patients, often struggling with health challenges, the possibility for exploitation or coercive manipulation leading to uninformed data transactions is highly probable. This is caused by the inherent complexity of providing health-related information and the possibility for individuals to consent to data trade without fully comprehending the far-reaching implications.

Second, the ethical issue extends to individuals with disabilities, who sometimes might encounter diminished cognitive capacities, or an inability to effectively articulate their preferences, especially if they are unable to speak. Such individuals are at an increased risk of being subject to undue influence or coercion, necessitating heightened safeguards to protect their interests.

Third, considerations of social class differences come into play, where individuals with financial precarity might be more susceptible to exploitation. This economic vulnerability could drive some to sell their health data in greater volume, or to undervalued prices. Such disparities are exacerbated in regions that have historically been subject to economic exploitation by wealthier counterparts. Without depicting extreme scenarios as in Boris Kunz's "Paradise" movie (2023), we must nevertheless consider that these risks are actual and real.

Additionally, the digital divide presents several challenges. Research has shown that the users of health technologies are mostly young and wealthy

individuals.[10] Individuals lacking access to the infrastructure or digital literacy may be excluded from participating in data marketplaces. This further perpetuates existing societal disparities, and potentially marginalises those who are technologically impaired.

However, the most concerning problem is what happens to health data once it is anonymised and out of the patient's control. This article will show the shortcomings of the concept of "anonymised data", and recent cases of data exploitation, which are inadequately addressed by the current legal framework. DataCom explores the possibility of changing the current regulations, taking into consideration the abovementioned issues.

II. State of the art

The scholarly debate surrounding data commodification is characterised by divergent perspectives on the treatment of personal data as a tradeable commodity.[11]

On one side of this debate are scholars who argue against the tradeability of personal data, asserting that data protection is an inalienable fundamental right guaranteed by Article 8 of the Charter of Fundamental Rights.[12] According to this viewpoint, data protection stands as a fundamental principle safeguarding individuals' rights, which cannot be waived or compromised by data subjects.[13] In this view, the status of personal data as a fundamental right elevates it above economic transactions, placing it beyond the reach of commercial exploitation.[14] This perspective argues that personal data cannot be considered a commodity to be traded. Therefore, it is not possible for private companies to exploit it, making the health data marketplace stand against the fundamental principles of the EU.

---

[10] Quinn Grundy, 'A review of the quality and impact of mobile health apps' (2022) 43 Annual review of public health 117.

[11] Vincenzo Ricciuto, 'La patrimonializzazione dei dati personali. Contratto e mercato nella ricostruzione del fenomeno' in Vincenzo Cuffaro, Roberto D'Orazio, Vincenzo Ricciuto (eds.), *I dati personali nel diritto europeo* (Torino 2019).

[12] EDPS – EDPB, 'EDPB-EDPS Joint Opinion 2/2022 on the Proposal of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)' [2022] EDPB/EDPS, Brussels.

[13] Bart Custers and Gianclaudio Malgieri, 'Priceless data: why the EU fundamental right to data protection is at odds with trade in personal data' (2022) 45 Computer Law & Security Review 105683.

[14] Svetlana Yakovleva and Kristina Irion, 'Toward Compatibility of the EU Trade Policy with the General Data Protection Regulation' (2020) 114 American Journal of International Law 10.

An alternative perspective asserts that the economic value inherent in personal data can be exploited while upholding the principles of data protection.[15] This viewpoint builds on the premise that data protection, while significant, should not be an absolute right. Rather, it must be balanced with other principles, such as business freedom.[16] This perspective holds that personal data can be treated as an economic asset, albeit with certain limitations and under specific circumstances.[17] The legal basis of consent, as stipulated in Article 6 GDPR, is often invoked in this context as a solution. According to this perspective, data subjects can provide informed consent for using and sharing their personal data, allowing for a level of data commodification that aligns with data protection laws and regulations.[18] This approach reconciles the potential economic benefits of personal data with the imperative to protect individuals' privacy rights.[19]

The framework introduced by the DSC might pose some problems, as highlighted by several opinions of the EDPB. Legal scholars have also been concerned with the potential conflicts between the DSC and existing data protection regulations, such as the GDPR.[20] However, limited attention has been devoted to investigating how these proposals may exacerbate the commodification of health data.

1. The EU Digital Strategy

The ambitious Digital Strategy Corpus of law (DSC) is torn between these polarized viewpoints, aiming to construct a harmonised digital market while safeguarding citizens' rights[21]. It consists of a series of legal instruments, including the Open Data Directive, the Data Governance Act (hereinafter:

[15] Beate Rössler, 'Should personal data be a tradable good? On the moral limits of markets in privacy' in Beate Rössler and Dorota Mokrosinska (eds.), *Social dimensions of privacy: Interdisciplinary perspectives* (Cambridge 2015) 141.

[16] Milena Mursia and Carmine Andrea Trovato, 'The commodification of our digital identity: limits on monetizing personal data in the European context' (2021) 2 Media Laws 165.

[17] Francesca Ferretti, 'Directive (EU) 2019/770: personal data as consideration in contracts for the supply of digital content and digital services and the inherent impact on privacy law' (2021) Actualidad Jurídica Iberoamericana 16, 1740–1777.

[18] Dianora Poletti, 'Le condizioni di liceità del trattamento dei dati personali' [2019] Giurisprudenza italiana 2783.

[19] Roberto Senigaglia, 'La dimensione patrimoniale del diritto alla protezione dei dati personali' (2020) 2(2) Contratto e impresa 760.

[20] Wenkai Li and Paul Quinn, 'The European Health Data Space: An expanded right to data portability?' (2024) 52 Computer Law & Security Review 105913.

[21] Carlo Botrugno, 'Cybersecurity, privacy and health data protection in the digital strategy of the European Union' (2022) 14(3) Rechtd. Revista De Estudos Constitucionais, Hermenêutica e Teoria Do Direito 300.

DGA), the AIA, the EHDS, and the Data Act. This legislative effort aims to enhance the responsible reuse of data while protecting individual rights. Yet, these provisions create several challenges and legal conflicts, often arising from disparities between the DSC's provisions and the existing data protection framework.[22]

The introduction of new provisions within the DSC, has been an important step in the EU digital legal landscape regarding data sharing and data reuse: several new mechanisms are provided by the EHDS,[23] the DGA,[24] the DA,[25] and the AIA. While surely driven by justified public interests, such as improving research and providing better healthcare services,[26] these provisions may endanger EU citizens, failing to find a balance between collective well-being and individual rights. This increase in data-sharing opportunities carries several risks, especially when in the hands of foreign entities.

The legal doctrine, for example, has raised several concerns regarding the reuse mechanism provided by the EHDS, such as the fact that it might diminish the possibility for patients to exert control over their data.[27] The ability of private companies, such as pharmaceutical corporations, insurance providers and banks, to exploit large datasets to make decisions that may prevent access to healthcare, especially in systems tied to insurance mechanisms, requires a deeper evaluation by legal scholarship.[28] In this context, the urgency of an academic analysis assumes a critical role in exploring the dangers of data commodification, considering the protection of special categories of data and ethical data sharing. The existing debate falls short of comprehending the extent to which these initiatives may exacerbate the commodification of health data in the private and public sectors.

The DA, on the other hand, constitutes a paradigmatic case aimed at facilitating data exchange from Internet of Things (hereinafter: IoT) devices

[22] Mahsa Shabani, 'Will the European Health Data Space change data sharing rules?' (2022) 375(6587) Science 1357.

[23] For example, the EHDS extends the application of the right to data portability to 'all electronic health data, including inferred data'. See Wenkai Li and Paul Quinn, 'The European Health Data Space: An expanded right to data portability?' (2024) 52 Computer Law & Security Review 105913.

[24] One novelty is the creation of a "Data Altruism", a concept that regards data shared for non-commercial use, without expecting a financial reward.

[25] The DA, for example, mandates data sharing by companies with public sector entities in exceptional situations, such as pandemics.

[26] Tjasa Petrocnik, 'Health data between improving health (care) and fueling the data economy' (2022) 2022 Technology and Regulation 124.

[27] Luca Marelli *et al* 'The European health data space: Too big to succeed?' (2023) 135 Health policy 104861.

[28] In principles, the EHDS explicitly forbid the detrimental use of health data, but fails to provide adequate safeguards to ensure that such provisions are respected.

between companies and users. A salient feature of this legislation lies in its capacity to promote the portability and interoperability of usage data originating from IoT devices, a substantial proportion of which frequently contain a spectrum of health-related information, such as those derived from tele-health devices, often a mixture of personal and non-personal data. By improving data sharing from IoT devices, the Data Act creates opportunities for data aggregation and subsequent analysis by small companies and individuals.[29] However, it might contribute to the trafficking and exploitation of health data, not only related to patients, but also to oblivious employees or private individuals.

The provisions of the AIA also raise several concerns.[30] In particular, it allows companies to acquire and store special categories of data for the purpose of a bias mitigation and testing,[31] and establishes a legal basis (*i.e.*, a legal provision) according to the GDPR. Companies such as Google, Apple and other technology providers now have the perfect pretext to use health data collected from users (*e.g.*, through smartphones) to train their algorithms (as they often write, "improve the service"). Given these issues, the AIA acquires a central position within the scholarly debate around health data commodification, and, therefore, its implications should receive more scholarly attention.

III. The concepts of anonymised data and the problem of re-identification in literature and case law

In recent years, a growing number of academic works have been dedicated to studying anonymisation techniques, and the minimum conditions required for data to be considered "anonymised", instead of "pseudonymised".[32] According to one interpretation, as long as the original data set exists, it will always be possible to re-identify the data. According to another viewpoint,

[29] Chiara Gallese-Nobile, 'A first commentary to the proposal for a new Regulation on fair access and use of data (Data Act)' (2022) 3 Media Laws 237.

[30] I Glenn Cohen, Theodoros Evgeniou, Sara Gerke and Timo Minssen, 'The European artificial intelligence strategy: implications and challenges for digital health' (2020) 2(7) The Lancet Digital Health e376.

[31] Chiara Gallese, 'Predictive justice in light of the new AI Act proposal' [2022] available at SSRN <https://ssrn.com/abstract=4286023> accessed 21 May 2024.

[32] Emily M Weitzenboeck, Pierre Lison, Malgorzata Cyndecka and Malcolm Langford, 'The GDPR and unstructured data: is anonymization possible?' (2022) 12(3) International Data Privacy Law 184.

a risk-based approach must be followed, and data is considered anonymised if it can be re-identified with some effort.

Existing works on anonymisation have focused only on its various technical aspects; yet, the consequences of re-using anonymised health data sets on citizens' fundamental rights and freedoms have not generated much attention. The consequences and effects of employing different techniques for the anonymisation of health data, the risks connected to the reuse of anonymised data sets, the increasing risk of re-identification due to the linking of multiple data sets, and the availability of big data in the context of the public sector, have not yet been substantially investigated by legal scholarship.

According to the findings of the "Report on secondary use of health data through European case studies", conducted in view of the creation of the EHDS, researchers have identified a significant barrier to data sharing, caused by the lack of guidance on anonymisation at both national and international level.[33] This uncertainty around anonymisation processes has had several consequences. Within the EU, data users have adopted overly risk-averse behaviors, treating all data as personal data due to lack of clarity. This cautious approach, together with the lack of semantic interoperability and differing interpretations of key terms, has consequently hampered the reuse of health data in the EU, and it might be an obstacle for a smooth implementation of the European Health Data Space (EHDS).[34]

According to the Report, although Article 4 GDPR offers a definition, a similar lack of guidance exists regarding the pseudonymisation of health data. This has led to various approaches to pseudonymisation across European countries, and even within them, a lack of consensus on the degree of separation required between the re-identification key and the data user, for data to be categorised as pseudonymised.[35] The debate surrounding whether pseudonymisation or anonymisation is more appropriate in a specific case adds to this difficult scenario.

The utilisation of different pseudonymisation standards and methodologies, coupled with non-homogeneous data collection practices in the healthcare sector, has generated interoperability issues, proving to be a substantial obstacle to data sharing. In healthcare, data is often collected for primary purposes, such as delivering medical services or conducting clinical trials,

[33] Linda Abboud and others, *Report on secondary use of health data through European case studies* (Towards the European Health Data Space February 2022).

[34] Margarida Mateus, Maria Loureiro, A Raquel Fernandes, Miguel Oliveira and Ricardo Cruz-Correia, 'Implementation Status of the Proposal for a Regulation of the European Health Data Space in Portugal: Are We Ready for It?' (2023) 302 Studies in Health Technology and Informatics 48.

[35] Abboud and others (n 31).

without specific intentions of creating specific datasets that might be reused and employed in different systems (such as AI software). The data collection process is typically carried out by medical professionals who may lack IT expertise, leading to unstandardised and non-optimised data formats. In addition, the accuracy of the data collection in the context of healthcare has been challenged: pilot studies have shown that, in some cases, administrative data presents an error rate of over sixty percent.[36] Consequently, the data may be insufficient or biased when repurposed for reuse, such as AI training, compromising the fairness and accuracy of the resulting AI models.

This can result in elevated costs when third-party involvement is necessary to align data sets, or when the implementation of additional non-standard safeguards is requested. The divergence in pseudonymisation practices among various entities further exacerbates the interoperability challenges, preventing collaboration in the data-sharing ecosystem, and thus compromising an effective implementation of the EHDS.

1. The concept of anonymisation in recent case law

A recent ruling by the EU General Court[37] has introduced nuanced distinctions between pseudonymous and anonymous data. This ruling emphasises that supervisory authorities must thoroughly evaluate whether data can be classified as personal based on subjective criteria.[38] According to this ruling, pseudonymised data might not always be considered personal data, creating a situation where the same dataset can be categorised differently based on the capacity of each entity to identify the data subjects. This notion seems to contrast Recital 26 GDPR[39] and the previous Patrick Breyer case.[40]

The context of the judgment stems from claims made against the Single Resolution Board (hereinafter: SRB), by the European Data Protection Supervisor (hereinafter: EDPS), in relation to data sharing within a resolution

[36] Angelo Chiappetta, *Proposta di studio clinico sugli esiti della colecistectomia nel Veneto,* Presentation at the Best specialization dissertation Prize event, Società Triveneta di Chirurgia, 21/12/2018 2022 – courtesy of the author.

[37] Case T-557/20, *Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS)* [2023], ECLI:EU:T:2023:219.

[38] Gonzalo F Gallego, Santiago De Ampuero Castellanos and Juan Ramon Robles, 'Sending personal data, receiving non-personal data: Recent EU judgment reinforces the power of pseudonymization' [2023] Lexology.

[39] Michèle Finck and Frank Pallas, 'They who must not be identified – distinguishing personal from non-personal data under the GDPR' (2020) 10(1) International Data Privacy Law 11.

[40] Case C-582/14 *Patrick Breyer v Bundesrepublik Deutschland* [2016] EU:C:2016:779.

scheme of a Spanish Bank. The SRB shared data after a seudonymizedon process, prompting a dispute on whether the shared data should be regarded as personal data. The EDPS argued that seudonymized data should be treated as personal data.[41]

However, the General Court annulled the EDPS's decision, highlighting the necessity to assess the recipient's capability to re-identify individuals from pseudonymised data: "Therefore, since the EDPS did not investigate whether Deloitte had legal means available to it which could in practice enable it to access the additional information necessary to re-identify the authors of the comments, the EDPS could not conclude that the information transmitted to Deloitte constituted information relating to an 'identifiable natural person' within the meaning of Article 3(1) Regulation 2018/1725".[42]

The distinction between pseudonymity and anonymity has always been complex, with technical advancements making "true anonymisation" challenging, to the point that several scholars and professionals believe it is impossible to anonymise health data.[43] Authorities often consider datasets containing data that could be linked to individuals with the help of third parties, as containing personal data. Before the recent ruling, the Court of Justice of the European Union (hereinafter: CJEU) established a very high threshold for anonymity, where even dynamic IP addresses could be identified with additional information, considering them personal data.

This ruling, however, clarifies that the determination of personal data should be based on the position and powers of each party, allowing different categorisations of the same data in different hands. This implies that data-sharing assessments must consider the recipient's perspective and their ability to re-identify data subjects.

The establishment of clear and harmonised guidelines at both national and international levels is, thus, crucial to facilitate responsible and effective data-sharing practices.[44] However, in the healthcare context, this lack of guidance is even more concerning because anonymised health data are outside the scope of the GDPR.

On the 9th November 2023, in the *Scania* case, the CJEU examined whether a vehicle's VIN number could fall within the concept of "personal data",

[41] Gonzalo and others (n 35).

[42] Case T-557/20 *Single Resolution Board (SRB) v European Data Protection Supervisor (EDPS)* [2023], ECLI:EU:T:2023:219, par. 105.

[43] Luc Rocher, Julien M Hendrickx and Yves-Alexandre De Montjoye, 'Estimating the success of re-identifications in incomplete datasets using generative models (2019) 10(1) Nature communications 1.

[44] Lubna Luxmi Dhirani, Noorain Mukhtiar, Bhawani Shankar Chowdhry and Thomas Newe, 'Ethical dilemmas and privacy issues in emerging technologies: a review' (2023) 23(3) Sensors 1151.

within the meaning of Article 4(1) GDPR.[45] VIN is defined by Article 2(2) of Regulation No 19/2011 as an alphanumeric code assigned to a vehicle by its manufacturer in order to ensure the proper identification of that specific vehicle. The VIN number is more detailed than a plate number: it is composed of the WMI (World Manufacturer Identifier: 3 alphanumeric identifying characters assigned to the manufacturer and its location), the VDS (Vehicle Descriptor Section: 6 alphanumeric characters describing the general characteristics of the vehicle), and the VIS (Vehicle Indicator Section: 8 characters, the first 4 of which are alphanumeric and the second 4 numeric, identifying the individual specimen).

The Advocate General believed that the VIN number is devoid of any "personal" character.[46] In their opinion, a VIN number acquires that character in relation to anyone who reasonably has the means to associate it with a specific person. The Advocate General also added that "where independent operators can reasonably be expected to have the means to link a VIN to an identified or identifiable natural person, which it is for the referring court to verify, that VIN constitutes personal data for them, within the meaning of Article 4(1) GDPR, as well as, indirectly, for the car manufacturers who make it available, even if the VIN is not, in itself, personal data for the latter and is not, in particular, when the vehicle to which that VIN has been assigned does not belong to a natural person".[47]

The CJEU first noted that the concept of personal data as "any information relating to an identified or identifiable natural person" is applicable where, "by reason of its content, purpose and effect, the information in question is linked to a particular natural person".[48] Subsequently, the CJEU held that "to determine whether a natural person is identifiable, directly or indirectly, account must be taken of all the means likely to be reasonably implemented either by the controller, within the meaning of Article 4(7) GDPR, or by others, to identify that person, without requiring that all the information enabling that person to be identified be in the hands of a single entity".[49]

Then, it concluded that "under point II.5 of Annex I to Directive 1999/37, the VIN must appear on the vehicle registration certificate, as must the name and address of the certificate holder. In addition, under points II.5 and II.6 of this annex, a natural person may be designated in the said certificate as the vehicle's owner or as a person who may dispose of the vehicle in a legal

[45] Case C-319/22, *Gesamtverband Autoteile-Handel v Scania CV AB*, ECLI:EU:C:2023:837.

[46] Case C-319/22, *Gesamtverband Autoteile-Handel v Scania CV AB*, ECLI:EU:C:2023:385, Opinion of AG Campos Sanchez-Bordona, paragraphs 34 and 39.

[47] Ibid.

[48] Ibid., paragraph 45.

[49] Ibid.

capacity other than that of the owner. In these circumstances, the VIN constitutes personal data, within the meaning of Article 4(1) GDPR, of the natural person mentioned in the same certificate, insofar as the person who has access to it could have means enabling him to use it to identify the owner of the vehicle to which it relates or the person who may dispose of that vehicle in a legal capacity other than that of the owner".[50]

The CJEU therefore adhered to the doctrine of the personalisation criteria, which considers personal data in relation to the subject who holds the data. It needs to be argued that this interpretation should be rejected, as it opens a dangerous precedent for health data and other sensitive information.

The qualification of personal data should be inherent to the data itself, regardless of the data processor, the data subject, or the data controller. Otherwise, data can be exploited and exposed to more risks, falling outside the scope of the GDPR.

For example, in Italy, the VIN number can be easily associated with the owner's personal information by simply paying 5 euros online through a quick VIM certificate request. The VIM certificate is a form in which official information can be retrieved by anyone from the ACI's PRA (Car Public Registry) database. By entering the VIM number, it is possible to verify a vehicle's technical data, its current owner (First and Last Name or Company Name, Tax Identification Number and/or VAT Number, Date of Birth, Residential Address or Registered Office), mortgages, current financing, leases, administrative detentions, theft verification, loss of possession, and verification of cancellation. There are similar services in all EU countries. Due to the widespread availability of these services, the personalisation criterion loses sense since the information is available to anyone with minimum effort.

The case of VIN numbers is just one of the many examples in which the digitalised and interconnected global society has made the retrieval of personal data a very easy task that any citizen could potentially perform. Under careful consideration of what personal data-based services are currently accessible online, it is visible that considering "personal" only data held by specific subjects is dangerous and anachronistic.

On the one hand, citizens tag themselves in every kind of social media or online service (*e.g.*, Google Maps), displaying their real name. They review websites, commercial activities, public services, and monuments worldwide. They post pictures with barcodes or metadata attached, revealing locations, exact dates, time stamps, and other kinds of information. On the other hand, transparency obligations and digitalised public services release personal data to the public. Anyone can put all this information together without any effort

[50] Ibid., paragraphs 47 and 48.

or with minimum effort (*e.g.*, relying on a paid service). Then, it becomes easy to link such information with health datasets, for examples those published after a ransomware attack.

However, several risks and adverse outcomes might materialise if health datasets are not adequately protected, both from a legal and a factual point of view, and this includes the qualification of "personal" or "anonymised" data. What would happen if the information in the electronic health records were publicly shared, arguing that only hospitals or doctors can identify the patients? This scenario is not impossible after the abovementioned rulings.

As early as in the 1990s, researchers warned about re-identification risks. Studies where computer scientists have effectively demonstrated the susceptibility of anonymised datasets to unauthorised access and re-identification have been well-documented.[51]

A notable example is the utilisation of such methodologies by Latanya Sweeney in 1997, wherein the identity of the then Massachusetts governor, William Weld, was successfully ascertained through a meticulous examination of publicly accessible hospital records.[52] Sweeney's approach primarily entailed a comparative analysis between anonymised hospital data pertaining to state personnel, and the voter registration records for the city of Cambridge. Given her knowledge of the governor's residence in Cambridge, the researcher cross-referenced the aforementioned datasets. This procedure enabled her to pinpoint specific records, discernible based on demographic factors such as age and gender, which corresponded solely to William Weld. Consequently, she showed that these records provided insights into Governor Weld's recent medical sojourn, including his medical diagnosis and the specific pharmaceuticals prescribed during the hospital visit, even though the dataset had been anonymised.

Previous opinions from the Privacy Authorities, such as the Irish DPA and the UK' Information Commissioner's Office (ICO), seemed to be very aware of these risks, and followed the approach whereby anonymisation should account for the possibility for third parties to identify data subjects.[53]

---

[51] Arvind Narayanan and Vitaly Shmatikov, 'Myths and Fallacies of Personally Identifiable Information' (2010) 53 Communications of the ACM 24, 26; Rocher and others (n 39); Arvind Narayanan and Vitaly Shmatikov, 'Robust De-anonymization of Large Sparse Datasets' in *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, (Oakland 2008) 111–125.

[52] Latanya Sweeney, 'Weaving technology and policy together to maintain confidentiality' (1997) 25(2–3) The Journal of Law, Medicine & Ethics 98.

[53] Finck (n 36).

2. The concept of anonymisation in technical literature

Technical literature in computer science and engineering on anonymisation techniques has shown that there is no conflict between the concept of anonymised data and that of non-anonymised data; anonymisation is not a black-or-white concept but rather a spectrum of grey shades. Different degrees of anonymisation can be used depending on how much data utility is needed to be preserved. In fact, as anonymisation becomes more robust, the utility of the data generally decreases, creating a trade-off situation that needs an assessment based on the objectives of the data use. Therefore, it is argued here that the traditional dichotomy between anonymised data and personal data should be reconsidered.

There are several different techniques that are insufficient – by themselves alone – to guarantee that the absolute re-identification risk will not arise:[54]

- Pseudonymisation: While not strictly an anonymisation technique, pseudonymisation replaces identifiers with fictitious labels (pseudonyms). This can be reversed when the key to link pseudonyms with original identifiers is available. Hence, it is often considered a weaker form of data protection.
- K-anonymity: This technique ensures that individual records cannot be distinguished from at least $k-1$ other records regarding certain "identifying" attributes. It aims to prevent "re-identification" of individuals by making each record indistinguishable from others in the dataset.
- L-diversity: An extension of k-anonymity, l-diversity requires that for each group of records sharing a combination of key identifying attributes, there must be at least 'l' distinct values for sensitive attributes. This method helps to prevent attribute disclosure.
- T-closeness: This method further refines l-diversity by requiring that the distribution of a sensitive attribute in any anonymised group is close to the distribution of the attribute in the overall dataset, typically within a 't' threshold. This approach helps in maintaining the "closeness" of attribute values, which protects against certain types of inference attacks.
- Differential Privacy: This is a more sophisticated method that provides a mathematical guarantee of privacy. It introduces randomness into the data aggregation process, such that the impact of including or excluding a single database record is minimal, thus masking the presence or absence of any individual.

[54] See Article 29 Data Protection Working Party, "Opinion 05/2014 on Anonymisation Techniques" (2014), 0829/14/EN WP216.

- Data Masking and Perturbation: These techniques modify, scramble, or obfuscate data to prevent direct linkage to identities. Methods include randomisation, data swapping, or noise addition. The challenge is to modify the data enough to prevent re-identification while still preserving its utility for analysis.

Each of these methods offers different degrees of protection, and affects the utility of the data in varying ways. The choice of method often depends on the specific context of use, the sensitivity of the data, and the required level of privacy.

For example, let's imagine a case in which we want to train a neural network on patients' data, in order to make a precise diagnosis. We have a hospital dataset that contains the following information: Name; Surname; Date of Birth; Sex of birth; Nationality; Ethnicity; Address; ZIP code; City; Country; Insurance; Income for reimbursement; Medication taken; Pre-existing diseases; Dates of admittance and dismissal from hospital; Surgery undertaken; Family history of diseases; Genetic information; Blood analysis values; BMI; Height; Weight; Physiological values; EEG; ECG. All of these individual pieces of information can be used to identify the patients even when we take out their names, since it is possible to match them with existing datasets (from voters' lists, social insurance lists, taxpayers databases, bank or health insurance databases, social media, information stored by private services providers, etc.). The more information we keep on our server, the greater the risk of re-identification. Therefore, we need to apply anonymisation techniques to protect the data.

In Case A, entries that are of no use for a medical diagnosis, such as name and surname, nationality, income, and country (making the data "pseudonymised" according to the GDPR) could be deleted. Then, quasi-identifiers, such as the exact date of birth, the exact admittance dates, and the ZIP code could be masked, and thus include only the year of birth, the length of hospital stay, and the neighborhood. According to legal interpretation, the resulting data might be seen as anonymised. However, it is necessary to be aware that the re-identification risk is still very high, due to the uniqueness of the medical history of each patient, and that linking it with other sources would still be very easy.[55] The data, however, would make it possible to have a very precise model, which could find useful connections between data (such as the prevalence of a certain disease risk in a given neighborhood).

In Case B, it could be decided to apply an additional layer of anonymisation techniques, which involves the masking of exact data with broader information, such as a slightly different year of birth, broad residence location, a general

[55] As noted by Michael Veale, Reuben Binns and Jef Ausloos, "When Data Protection by Design and Data Subject Rights Clash" (2018) 8 International Data Privacy Law 105, 107.

category of pre-existing disease, or a broad idea of ethnicity (*e.g.*, instead of African American, we could write "African descendant", which includes African people living in the EU). Instead of the whole genomic information, only a specific piece that is relevant to the given analysis could be included. Those transformations would make the data "more anonymised", but still very useful for our model.

In case C, however, when all the remaining information is not needed, differential privacy could be applied, or other techniques to mask the data further, such as providing a range of ages instead of a year, making the dataset "even more anonymised", but still useful to train AI to a certain degree.

If anonymisation would go even further, like aggregating the data, the model cannot be trained anymore, as the information needed is lost, but the level of data protection is extremely high. When deleting the original raw data, and there is no way to trace the source and the identity of the patients in any way, then the risks for them are at their minimum.

IV. Recent cases of health data commodification and common risks

The topic of health data commodification, and the implications of the Digital Strategy Corpus of law (DSC), necessitates a broad analysis that considers not only the legal and ethical dimensions on data protection, but also the technical considerations explored in the previous section, taking into account the geopolitical context. Recent developments highlight the potential vulnerabilities of health data sharing, especially when it transcends national borders, and involves partnerships with foreign entities that might have a very different legal system and an opposed data culture or "health data consciousness".[56]

Once the data is transferred to these countries, data subjects most likely lose control over it for good, as it is often too complex to have access to legal assistance abroad (assuming that data subjects are aware of the international transfer, which might not always be the case).

The concerns raised by the U.S. Government, regarding the potential exploitation of health data by foreign companies,[57] can be seen as a valid

[56] "Health data consciousness" means here a culture surrounding health data protection, which includes cultural, institutional, and personal factors such as ideas, views, feelings, and traditions. The term is inspired by the concept of "legal consciousness" Susan S. Silbey, 'After legal consciousness' (2005) 1 Annu. Rev. Law Soc. Sci. 323–368.

[57] Mark Kazmierczak and others, *China's Biotechnology Development: The Role of US and Other Foreign Engagement* (US-China Economic and Security Review Commission 2019).

example for the EU, especially considering the increasing geopolitical tensions, and the proliferation of malicious apps and emails aimed at trafficking users' data. Digital connections facilitate cross-border movement of data, often without adequate oversight and regulation. In this context, the DSC's might learn from the challenges faced in the U.S. healthcare industry.

The DSC provisions, originally intended to enhance data sharing within the EU market, could expose health data to external parties without adequate safeguards, especially when it is considered to be anonymised data, and thus out of the GDPR protection. The recent adequacy decision regarding data transfer between the EU and the US[58] further complicates the scenario for both regions.

For example, the proliferation of digital health apps, designed to provide users with insights into their health, has raised concerns about the potential exploitation of health information, particularly in the context of pregnancy apps,[59] reproductive health applications, or mental health support apps,[60] where individuals share highly personal health data, often without a full understanding of how that data might be used, and with whom it might be shared.

1. The case of smartphone applications

Smartphone apps are an ideal case study as they have gained widespread popularity,[61] providing women with tools to track their fertility cycles, access reproductive health information, monitor their pregnancy journeys, share views and expectations with peers, and other sensitive information (*e.g.*, ultrasound scans). However, this increase in popularity has also exposed vulnerabilities in data protection measures. Many pregnancy apps are developed by companies based in the US, or other foreign countries, raising questions about the compatibility of data collection, data storing, and sharing practices by these companies with the GDPR and other EU regulations (such as the Medical

[58] See European Commission, 'Data Protection European Commission adopts new adequacy decision for safe and trusted EU-US data flows' (10 July 2023) <https://ec.europa.eu/commission/presscorner/detail/en/ip_23_3721> accessed 30 September 2023.

[59] Jo-anne Patricia Hughson, Oliver Daly, Robyn Woodward-Kron and John Hajek, 'The rise of pregnancy apps and the implications for culturally and linguistically diverse women: narrative review' (2018) 6(11) JMIR mHealth and uHealth e189.

[60] Chiara Gallese, 'Legal Issues of the Use of Chatbot Apps for Mental Health Support', in Alfonso González-Briones and others (eds.) *Highlights in Practical Applications of Agents, Multi-Agent Systems, and Complex Systems Simulation. The PAAMS Collection*, (Berlin 2022) 258.

[61] Yeonkyu Lee and Mikyung Moon, 'Utilization and content evaluation of mobile applications for pregnancy, birth, and child care' (2016) 22(2) Healthcare informatics research 73.

Device Regulation). Many app developers have not demonstrated adequate safeguards to protect user data,[62] leading to potential violations of data protection regulations.

A significant risk is sharing personal health data with third parties without user consent, which is a problem shared with health platforms.[63] Research by Privacy International has revealed that some health apps share deeply personal and sensitive data with third-party entities, including companies like Facebook.[64] This practice is particularly alarming when considering the special character of health data, which requires enhanced protection due to its sensitive nature. Despite these legal requirements, app developers' lack of data protection measures leaves users vulnerable to data exploitation.

A case study of the UK maternity app Badgernet, introduced in 2017,[65] is also interesting. This app is part of the maternity care process, used by healthcare professionals and pregnant women. It allows users to view medical notes, schedule appointments, and access information related to their pregnancy. However, the platform's handling of sensitive health information and its data protection measures has raised concerns in several ways, as highlighted by Privacy International.[66]

Firstly, the app's potential to hold significant amounts of personal information emphasises the need for robust data protection measures. Despite these requirements, the app lacks a data protection officer, potentially undermining compliance with legal standards. Moreover, Badgernet's privacy policy raises questions about the extent of data shared with third parties. Ambiguities in the policy obscure the nature and scope of data shared with entities such as Microsoft and Google. This lack of transparency has an impact on users' understanding of data processing, ultimately compromising their ability to make informed decisions about their own data.[67]

The case of Badgernet also highlights the potential risks of the digital divide phenomenon, excluding individuals without digital access from essential services. As the app becomes integrated into the maternity care service,

---

[62] Borja Martínez-Pérez, Isabel De La Torre-Díez and Miguel López-Coronado, 'Privacy and security in mobile health apps: a review and recommendations' (2015) 39 Journal of medical systems 1.

[63] Brígida Riso and others, 'Ethical sharing of health data in online platforms–which values should be considered?' (2017) 13 Life sciences, society and policy 1.

[64] Privacy International, *How Apps on Android Share Data with Facebook (even if you don't have a Facebook account)* (Privacy International 2018).

[65] Emilia Crighton, 'Public health screening programme annual report: 1 April 2018 to 31 March 2019' [2020] NHSGGC Public Health Directorate.

[66] Privacy International, *How digital health apps can exploit users' data* (Privacy International 2022).

[67] Privacy International, *How digital health apps can exploit users' data* (n 55).

those lacking digital devices or recent smartphones may face obstacles in accessing crucial healthcare information. The push towards digitalisation may inadvertently create disparities in healthcare access,[68] disadvantaging marginalised individuals who rely on alternative modes of communication.

The cases of health data exploitation by companies urge the implementation of ethical data governance principles within digital health. It is important that companies prioritise transparency, user consent, accountability, and data protection measures, in order to ensure that individuals are not subjected to data exploitation.[69]

Transparency is a key element in ensuring ethical data governance. App developers must provide users with clear and comprehensible information about how their data will be used, shared, and protected. Ambiguous privacy policies and data-sharing practices violate users' rights and erode user trust.[70]

Furthermore, the above cases highlight the broader issue of digital inclusion and access to technology. As technology becomes widespread in healthcare services, efforts should be made to ensure that marginalised communities are not left behind due to digital disparities.[71] Striking a balance between technological advancements and equitable access to healthcare technology is crucial.

2. Brokerage of mental health data

Another case study regards data brokerage and sensitive mental health information, unveiling a deeply concerning trend where personal data on individuals' mental health conditions is being marketed. In a previous study, some concerns regarding mental health apps due to data protection issues were highlighted.[72] A report by Joanne Kim sheds further light on the practices of

[68] Uchechi A Mitchell, Perla G Chebli, Laurie Ruggiero and Naoko Muramatsu, 'The digital divide in health-related technology use: The significance of race/ethnicity' (2019) 59(1) The Gerontologist 6; Chukwuma N Eruchalu and others, 'The expanding digital divide: digital health access inequities during the COVID-19 pandemic in New York City' (2021) 98 Journal of Urban Health 183.

[69] Mackenzie Graham, 'Data for sale: trust, confidence and sharing health data with commercial companies' (2023) 49(7) Journal of Medical Ethics 515.

[70] Urs-Vito Albrecht, 'Transparency of health-apps for trust and decision making' (2013) 15(12) Journal of Medical Internet Research e277.

[71] William J Gordon, Adam Landman, Haipeng Zhang and David W. Bates, 'Beyond validation: getting health apps into clinical practice' (2020) 3(1) NPJ Digital Medicine 14.

[72] Gallese, 'Legal Issues of the Use of Chatbot Apps for Mental Health Support' (n 49).

data brokers in the acquisition and sale of mental health data,[73] revealing that some data brokers are actively and unethically marketing mental health data. Of the 37 data brokers contacted, 26 responded to inquiries about mental health data, and 11 were willing to sell such data. This practice is alarming, as it involves the trade of highly personal information that can have significant implications for individuals' mental well-being and privacy, especially if leaked to employers, due to the social stigma still attached to mental health conditions.

The research by Kim reveals that the data sold by data brokers includes information on individuals with conditions such as depression, attention disorders, insomnia, anxiety, ADHD, and bipolar disorder. Additionally, they provide data on ethnicity, age, gender, zip code, religion, marital status, net worth, credit score, date of birth, and single parent status.[74] This aggregation and subsequent sharing of personal and mental health information raises concerns about the potential misuse of such data for profiling and discriminatory practices.

The report also highlights a lack of stringent vetting procedures and regulation on the use of purchased data. Many data brokers are not clear on whether the data will be de-identified or aggregated, potentially leaving room for the sale of identifiable data. In addition, according to the report, despite initial inquiries about the purpose and cases of use of the data, data brokers do not appear to have additional controls for client management, or background checks, to corroborate clients' statements. Pricing for mental health information varied widely among data brokers. While some charged relatively lower amounts, others demanded substantial fees for a subscription or licensing access to data.[75]

These cases raise many legal and ethical issues about the practices of data brokers. The lack of clear consent mechanisms, and the potential for data to be used without individuals' knowledge or explicit permission, undermines the principles of informed consent, without mentioning all other GDPR principles that often are not even considered. Individuals may be unaware of how their sensitive mental health information is being bought, sold, and potentially exploited by third parties.

---

[73] Joanne Kim, *Data Brokers and the Sale of Americans' Mental Health Data* (Duke Sanfort 2023).

[74] Ibid.

[75] Ibid.

The sale of mental health data without adequate safeguards can contribute to stigma[76] and discrimination against those with mental health conditions.[77] Moreover, the potential misuse of this data for targeted marketing and other purposes could exacerbate individuals' mental health struggles and compromise their well-being.

Although the report specifically explores the US situation, it is also relevant for EU citizens because most apps are developed by US companies, which can easily process EU data due to the recent adequacy decision.

The report is, however, just one example of the risks of health data exploitation in the private sector, which calls for strict regulation of the data brokerage industry, particularly concerning sensitive data such as mental health information. The DSC should address issues related to data anonymisation and data use limitations, mandating a human rights assessment and a risk management system. Additionally, data brokers (even when they are foreign entities) should be forced to adopt transparent practices that provide individuals with the means to understand the implications of data processing, beyond the information required by the GDPR.

It is also important to empower users to have control over their data.[78] Users should be informed about the potential risks associated with sharing their data. They should have the ability to make informed decisions about whether, and how, their data is processed by data brokers, even if de-identified.

The exploitation of mental health data by data brokers is a matter that concerns society at large: with a growing population of elderly people, the risks of exploitation and the lack of informed consent are imminent. Stakeholders should establish clear guidelines and ethical standards, involving policymakers, regulators, healthcare providers, and technology companies.

Another similar case is the sale of health data of rape victims, which data brokers also targeted, as testified by Pam Dixon in 2013 at a Senate hearing. In her testimony, Dixon said that a pharmaceutical marketing company had advertised a list of a thousand rape victims at the price of 79 dollars.[79] This

[76] Graham Thornicroft and others, 'The Lancet Commission on ending stigma and discrimination in mental health' (2022) 400(10361) The Lancet 1438.

[77] Public Interest Advocacy Centre, *Mental health discrimination in insurance* (Public Interest Advocacy Centre, Liverpool, Sydney 2021).

[78] Katherine Elmlinger, 'My Data, My Choice? An Analysis of the Data Broker Industry and the Exercise of Privacy Choices' [2023] Princeton University Undergraduate Senior Theses.

[79] For further reading regarding the testimony and the problem of health data brokers in the US, see the U.S. Government Publishing Office, 'What information do Data Brokers Have on Consumers, and how they use it?', (2015) <https://www.govinfo.gov/content/pkg/CHRG-113shrg95838/html/CHRG-113shrg95838.htm> accessed 30 September 2023.

is just one of the many cases where data subjects are being exploited by data brokers[80].

Even assuming that, in the above-mentioned cases, data subjects somehow consented to the sale of their data (as known, only a few individuals read privacy policies and T&C of online services), ethical concerns remain. The US situation can be regarded as an example of what could happen in the EU, if the matter is not adequately regulated. It is among the core EU values to protect vulnerable groups, and freedom of doing business should not be used as a tool to exploit their data: for this reason, the market under the Digital Strategy should be closely supervised.

V. Health data reuse under the European Health Data Space and the AI Act

Both the European Health Data Space and the AIA provide new mechanisms for re-using health data that require deeper analysis, given the considerations illustrated in the previous sections.

1. The European Health Data Space (EHDS)

Section IV of the EHDS is dedicated to the secondary use of electronic health data[81]. Article 34 illustrates the purposes for which data can be shared:

[80] See <https://www.commerce.senate.gov/2013/12/what-information-do-data-brokers-have-on-consumers-and-how-do-they-use-it> accessed 30 September 2023.

[81] Article 33 "Health data holders shall make the following categories of electronic data available for secondary use in accordance with the provisions of this Chapter: (a) electronic health data from EHRs; (b) data on factors impacting on health, including socio-economic, environmental and behavioural determinants of health; (ba) aggregated data on healthcare needs, resources allocated to healthcare, the provision of and access to healthcare, healthcare expenditure and financing; (c) pathogen data, impacting on human health; (d) healthcare-related administrative data, including dispensation, claims and reimbursement data; (e) human genetic, epigenomic and genomic data; (ea) other human molecular data such as proteomic transcriptomic, metabolomic, lipidomic and other omic data; (f) automatically generated personal electronic health data, through medical devices; (fa) data from wellness applications; (g) data on professional status, specialisation and institution of health professionals involved in the treatment of a natural person; (h) population-based health data registries (public health registries); (i) data from medical registries and mortality registries; (j) data from clinical trials, clinical studies and clinical investigations subject to Regulation (EU) 536/2014, Regulation (EU) 2017/745 and Regulation (EU) 2017/746, respectively; (k) other health data from medical devices; (ka) data from registries for medicinal products and medical devices; (l) data from

"(a) public interest in the area of public and occupational health[82] […];
(b) policy making and regulatory activities to support public sector bodies or Union institutions, agencies and bodies […];
(c) statistics […] related to health or care sectors […];
(d) education or teaching activities in health or care sectors at the level of vocational or higher education;
(e) scientific research related to health or care sectors, contributing to public health or health technology assessment, or ensuring high levels of quality and safety of health care, of medicinal products or of medical devices, with the aim of benefitting the end-users, such as patients, health professionals and health administrators[83] […]".

Individuals have the right to opt-out of having their personal electronic health data made available for secondary use, with mechanisms for opting out provided clearly and accessibly. However, exceptions are made for public interest purposes, such as health protection or significant research (see the paragraph related to Italian law), where Member States can establish mechanisms to access the data of those who have opted out under certain conditions. In addition, the opt-out mechanism is voided by Article 48 EHDS when data are de-identified: "If the purposes for which a health data holder processes personal electronic health data do not or do not longer require the identification of a data subject by the controller, the health data holder shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with the right to opt out under this Article". Therefore, the provision essentially permits the use of de-identified data even when the data subject has opted out.

To strengthen data protection further, and to clarify roles and responsibilities, it would be worth specifying guidelines or methods for securely de-identifying personal data, while ensuring that such measures are reversible in situations where re-identification becomes necessary under a different lawful basis. Additionally, it would be useful to include oversight mechanisms, in order to ensure that data holders or data users do not misuse this provision to impede the right of data subjects to opt-out when they want to.

There are also certain uses that are prohibited, as stated in Article 35 EHDS:

---

research cohorts, questionnaires and surveys related to health, after the first publication of results; (m) health data from biobanks and associated databases."

[82] "such as activities for protection against serious cross-border threats to health and public health surveillance or activities ensuring high levels of quality and safety of healthcare, including patient safety, and of medicinal products or medical devices".

[83] It includes "training, testing and evaluating of algorithms, including in medical devices, invitro diagnostic medical devices, AI systems and digital health applications".

"(a) taking decisions[84] detrimental to a natural person or a group of natural persons based on their electronic health data […];
(b) taking decisions in relation to a natural person or groups of natural persons in relation to job offers or offering less favourable terms in the provision of goods or services […] or taking any other decisions in relation to a natural person or groups of natural persons having the effect of discriminating on the basis of the health data obtained;
(c) advertising or marketing activities;
(e) developing products or services that may harm individuals, public health or societies at large […];
(eb) activities in conflict with ethical provisions pursuant to national law […]".

EU rules thus prohibit a detrimental secondary use, but not a "neutral" use (such as a for-profit activity where data subjects have no gain). In addition, the concept of "detrimental", as provided in the text, is not entirely clear, as it is not known if indirect damages (such as political decisions by third parties after the output of the research is published) are taken into consideration.

After having obtained such data, data users must publish their results or outputs, and "those results or output shall only contain anonymous data". However, as shown above, the concept of "anonymous" is very blurry and the EHDS does not provide any guidance on that matter.

2. The AI Act

Article 10, Paragraph 5 of the AI Act[85] provides for the conditions under which data controllers are allowed to reuse health data for de-biasing purposes.

[84] "in order to qualify as "decisions", they must produce legal, social or economic, effects or similarly significantly affect those natural persons".

[85] "To the extent that it is strictly necessary for the purpose of ensuring bias detection and correction in relation to the high-risk AI systems in accordance with paragraph (2), points (f) and (g) of this Article, the providers of such systems may exceptionally process special categories of personal data, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons. In addition to the provisions set out in Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680, all the following conditions must be met in order for such processing to occur: (a) the bias detection and correction cannot be effectively fulfilled by processing other data, including synthetic or anonymised data; (b) the special categories of personal data are subject to technical limitations on the re-use of the personal data, and state-of-the-art security and privacy-preserving measures, including pseudonymisation; (c) the special categories of personal data are subject to measures to ensure that the personal data processed are secured, protected, subject to suitable safeguards, including strict controls and documentation of the access, to avoid misuse and ensure that only authorised persons have access to those personal data with appropriate confidentiality obligations; (d) the special categories of personal data are not to be transmitted, transferred or otherwise accessed by other

Letter (a) emphasises the exclusivity and necessity of using special categories of personal data for bias detection, when alternatives (such as synthetic or anonymised data) are inadequate. It aligns with GDPR principles of data minimisation and specificity in data processing. However, it is necessary to clarify the criteria under which alternative data types are deemed inadequate for these purposes, in order to ensure a uniform application of this rule and avoid discretionary power of data controllers.

Letter (b) mandates robust technical and security measures to protect data integrity and confidentiality, while Letter (c) emphasises the enforcement of strict access controls, and the importance of documentation, which is essential for auditing and compliance purposes.

Letter (d) restricts data sharing, highlighting the privacy-first approach required in handling sensitive personal data. Letter (e) ensures data is not kept longer than necessary, adhering to GDPR principles of storage limitation. However, it does not specify mechanisms or criteria for determining when bias has been 'corrected' to guide data controllers and avoid misuse of data. Letter (f) also aligns with the GDPR, mandating recording of processing activities.

3. The Italian law on Artificial Intelligence

The Italian government went further in providing for a legal basis for the processing of special categories of data for scientific research. Article 9[86] of the approved bill is intended to complement the provisions of the EHDS.

---

parties; (e) the special categories of personal data are deleted once the bias has been corrected or the personal data has reached the end of its retention period, whichever comes first; (f) the records of processing activities pursuant to Regulations (EU) 2016/679 and (EU) 2018/1725 and Directive (EU) 2016/680 include the reasons why the processing of special categories of personal data was strictly necessary to detect and correct biases, and why that objective could not be achieved by processing other data".

[86] "1. The processings of data, including personal data, carried out by public and private non-profit entities for research and scientific experimentation for the creation of artificial intelligence systems for the purposes of prevention, diagnosis and treatment of diseases, development of drugs, therapies and rehabilitation technologies, creation of medical devices, including prostheses and interfaces between the body, and tools to support the patient's conditions, public health, personal safety, health and health safety, as necessary for the creation and use of datasets and basic models, are declared to be of significant public interest in implementation of article 32 of the Constitution and in compliance with the provisions of article 9 letter g) of EU Regulation 679/16. 2. For the same purposes, without prejudice to the interested party's obligation to provide information which can also be fulfilled by making general information available on the website of the data controller and without further consent of the interested party where initially required by law, the secondary use of personal data is always authorized, including those belonging to the categories indicated in Article 9 of EU

It details the conditions under which public and private non-profit entities can process personal data for research purposes in the field of AI related to health, safety, and the development of medical technologies. It qualifies such activities as falling automatically in the ambit of significant public interest. Declaring that these activities are of significant public interest, serves to justify the processing of sensitive personal data under specific GDPR provisions, specifically Article 9(g), which allows such processing for reasons of substantial public interest, and also exempts the data users from respecting the opt-out choices.

In addition, paragraph 2 of the Italian bill relaxes the requirements for obtaining consent for the secondary use of personal data in these contexts. It indicates that providing general information via the data controller's website can suffice.

Paragraph 3 of the Italian bill requires that any such processing receive approval from an ethics committee and be communicated to the Italian Data Protection Authority (DPA) with detailed compliance information. However, the DPA can intervene only within 30 days, and if no response is given, the activity is automatically approved. Nevertheless, paragraph 4 of the Italian bill reiterates the full authority of the DPA to inspect, block, and sanction, ensuring that these rights and powers are not diminished by the permissions granted in the earlier paragraphs.

VI. Does the DSC increase the risks of data commodification?

The previous sections introduced the Digital Strategy Corpus of law (DSC) and showed the risks of the exploitation of anonymised datasets. This section will explain why the current legal framework within the DSC is inadequate to protect citizens from the risk of having their data exploited. The solution to this problem is to include anonymised health data in the scope of the GDPR, and to introduce the concept of the "best interest of the patient."

Regulation no. 679/2016, by public and private non-profit entities. 3. The processing referred to in the previous paragraphs 1 and 2 must be approved by the ethics committees concerned and must be communicated to the Data Protection Authority with the indication of all the information required by articles 24, 25, 32 and 35 of EU Regulation 679/16 as well as the express indication, where present, of the data processors identified pursuant to article 28 of the aforementioned Regulation, and can be started thirty days after the aforementioned communication if not subject to a block ordered by the DPA. 4. All inspection, blocking, and sanctioning powers of the DPA remain intact".

The European Health Data Space (EHDS) provides for extensive exceptions to the opt-out possibility when certain conditions are met. However, it fails to consider the power imbalance between data subjects that are in a vulnerable position (*i.e.*, people with health conditions) versus public administration. In addition, as explained in the previous section, the excuse of de-identification can be easily exploited to prevent opting out. The opt-out mechanism was the only way in which patients retained control over their own health data and could protect themselves against abuses.

The AIA provision, establishing a novel legal basis for the processing of special categories of data, represents a backdoor to data misuse. Article 9 GDPR is very restrictive regarding the use of such data, but a simple modification of its discipline, such as the inclusion of the legitimate interest basis for de-biasing AI systems (with the same restrictions provided in Article 10 AIA), could have been sufficient to reach the same goal, with an appropriate degree of flexibility, without unnecessarily endangering patients' rights. Data Controllers would have been forced to perform the legitimate interest threefold assessment[87] instead of relying on the law *tout court*.

The legitimate interest test would make it possible to reflect on the necessity, balancing of rights, and appropriateness – with respect to the intended purpose of the data processing – and, most importantly, it would allow data subjects to object to data processing.

In addition, Article 10 AIA implicitly allows for the use of such data for training purposes. In fact, paragraph 6 specifies that "For the development of high-risk AI systems *not* using techniques involving the training of AI models, paragraphs 2 to 5 apply only to the testing data sets", implicitly saying that in other cases par. 2 to 5 apply to training and validation datasets. This creates a number of issues that the legislator fails to address, such as the fact that it would be impossible to respect paragraph 5, lett. e)[88] in all cases. Some models have been shown to retain the training data, and to disclose them in certain circumstances, and there are issues regarding accuracy.[89]

In the same way as for the reuse in the EHDS, this legal basis for reuse should be mitigated in view of patients' best interest.

---

[87] Article 29 WP, Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, 9 April 2014, 844/14/EN WP 217.

[88] "The special categories of personal data are deleted once the bias has been corrected or the personal data has reached the end of its retention period, whichever comes first".

[89] See ICO, "Generative AI third call for evidence: accuracy of training data and model outputs", ICO consultation series on generative AI, available at <https://ico.org.uk/about-the-ico/what-we-do/our-work-on-artificial-intelligence/generative-ai-third-call-for-evidence/> accessed 21 May 2024.

The major pitfall of the current legal framework is that both legal instruments do not prevent economic gain from patient data, or private companies from accessing the data. In addition, the transfer of data in third countries is not prohibited.

The EHDS even seems to be dismissive regarding the data minimisation principle, giving priority to the accuracy principle, as explained in Recital 39a.[90]

In addition, the fact that the EHDS provides for the possibility of granting access to data under the payment of a fee, is contradictory with the view of personal data as an untradeable commodity, because it can be seen as a sale, although reframed as a "contribution for incurred costs". Considering this scenario, there is a high risk of data commodification both at national and international level.

1. A proposal for a new legal discipline regarding anonymised health data

The three cases illustrated in paragraph III.2 are all examples of anonymisation according to the GDPR, but they have very different risks of re-identification. Moreover, the implications for patients' rights are not comparable. The less anonymised a dataset is, the more useful it is for the analysis, but at the same time, the more similar the risks for patients are to those of non-anonymised data. Due to the need for data utility and the provision of the AIA, the sharing mechanism in the European Health Data Space (EHDS) will likely enable the sharing of a large amount of loosely anonymised or pseudonymised data.[91] Needless to say, the increase in data sharing equals the increase in related data protection risks, such as cyber-attacks.

However, even the highest degree of anonymisation – such as data aggregation – is prone to affect citizens' rights. In fact, even in cases where insurance firms and financial institutions may encounter challenges in uniquely

[90] "The health data user who benefits from access to datasets provided under this Regulation could enrich the data with various corrections, annotations and other improvements, for instance by supplementing missing or incomplete data, thus improving the accuracy, completeness or quality of data in the dataset". Recital 5 also specifies that inferred data is included in the definition: "The electronic health data concern all categories of those data, irrespective to the fact that such data is provided by the data subject or other natural or legal persons, such as health professionals, or is processed in relation to a natural person's health or well-being and should also include inferred and derived data, such as diagnostics, tests and medical examinations, as well as data observed and recorded by automatic means".

[91] As stated in Article 44 "The health data access bodies shall provide electronic health data in an anonymised format, where the purpose of processing by the health data user can be achieved with such data, taking into account the information provided by the health data user".

discerning individual patients from aggregated datasets, their capacity to correlate health-related records with other diverse information data sources, such as a list of residential locations, remains a problem, especially with the help of AI models. These entities could potentially make detrimental decisions, such as the imposition of elevated premiums and financial rates within particular localities under the premise of heightened prevalence of specific diseases within said regions. Strategic utilisation of contextualised data allows these companies to direct their actions toward these specific demographic segments, notwithstanding their limited knowledge regarding the patients' personal identities.

In light of these considerations, regulating anonymised health datasets, in view of the EHDS and the AIA, becomes an urgent matter. Due to their nature, such data should be granted the same level of protection provided by the GDPR. Especially in the case of genomic data, head MRI scans, and other health data that are very difficult to anonymise, the GDPR could be a useful tool to protect citizens' data.

The traditional distinction between personal data and non-personal data is becoming outdated, in light of the advancement of technical research, which has shown that there is no such thing as "anonymised data" *tout court*, but there are only personal data processed with different techniques that preserve different degrees of information. In the case of health data, it is becoming particularly urgent to re-define the notion of personal data. As noted by Fink et al., "Accepting that there always remains a residual risk of identification even where data is anonymised appears to be the only realistic option in light of contemporary developments. Research has amply highlighted that anonymisation is never absolute. If the law were to insist that it must be, the only logical conclusion would be that data that once was personal data can only ever be pseudonymized but never anonymized. This would not only reject the spirit of Recital 26 GDPR in favour of an absolute approach but also radically change the nature and status of data protection law. Indeed, one would then need to rethink the very distinction between personal and non-personal data on which EU law is currently based". The seminal paper of Ohm[92] calls for the abandonment of the distinction between personal and non-personal data. It is argued here that this would be the ideal scenario in the case of health data. The fact that anonymised health datasets are considered personal data will not mean that they cannot be used anymore. They just must be treated with the same care as non-anonymised datasets.

If the scope of the GDPR is extended to include anonymised health data, however, clearer rules should be provided for scientific research, since many

[92] Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCL Law Review 1701; Michael Veale and others (n 46) 113.

companies are collaborating with research institutions and universities to develop diagnostic tools, and other products that need a large amount of health data to be trained. The AIA does not clarify this scenario.

In addition, the EHDS should be modified in order to adequately protect fundamental rights. The notion of "the best interest of the patient" should be incorporated into its discipline, mandating that a "best interest test" should be performed before health data reuse, inspired by family law, which builds on Article 3 of the UN Convention on the Rights of the Child. The same degree of protection should be granted to patients in all cases of data reuse, including research. If this principle is applied, the risks of misuse of anonymised datasets are greatly reduced. This principle should be incorporated into the AIA as well.

VII. Conclusion

As health data commodification increases, the EU's Digital Strategy Corpus of law (DSC) presents some important challenges related to health data, in particular anonymised datasets, which call for a redefinition of the concept of personal data. The digital, interconnected, and global world provides multiple occasions of health data sharing from patients to companies, posing major risks to fundamental rights. While the DSC has introduced various legislative measures aimed at facilitating data sharing, it is evident that there are still numerous risks and uncertainties that require thorough examination within the legal community.

The legal void regarding anonymised data is an issue of major concern, due to the lack of protection ensured for this type of data. Recent case law complicates this legal scenario further, giving rise to a dangerous interpretation of the concept of anonymisation.

Data brokers, which represent influential actors in the health data market, can currently freely exploit anonymised datasets, disregarding patient preferences. Given the EU's strong emphasis on upholding fundamental rights, it is anticipated that legislators will take proactive steps to address these pressing issues. Future regulations are thus likely to impose restrictions on the commercial use of anonymised health data, with the aim of safeguarding fundamental rights and freedoms.

To effectuate such a shift, a reform of data protection regulations is essential. Anonymised health data should be explicitly incorporated into the scope of the GDPR, as well as other relevant laws, ensuring that citizens are provided with additional safeguards and that their privacy remains protected from emergent practices of health data commodification. This transformation

in the legal framework will protect individual rights, but, at the same time, it will improve trust in the responsible processing of anonymised health data. In addition, such shift can promote the safer reuse of health datasets in important fields, such as scientific research, taking into consideration the best interest of the patients.

Funding

Project 101108151 – DataCom – HORIZON-MSCA-2022-PF-01. Funded by the European Union. Views and opinions expressed are however those of the author(s) only, and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the granting authority can be held responsible for them.

The cost of editing selected articles published in the Yearbook of Antitrust and Regulatory Studies in the 2022–2024 is covered by funding under the program "Development of scientific journals" of the Ministry of Education and Science under agreement No. RCN/SN/0324/2021/1. Task title: "Verification and correction of scientific articles and their abstracts". Funding value: 36 298,00 PLN; The task consists of professional editing of articles published in English.

Declaration of Conflict of interests

The author declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Declaration about the scope of AI utilisation

The author used standard AI tools to improve spellchecking, grammar and readability.

Literature

Abboud L and others, *Report on secondary use of health data through European case studies* (Joint Action Towards the European Health Data Space 2022).

Albrecht UV, 'Transparency of health-apps for trust and decision making' (2013) 15(12) Journal of medical Internet research e277.

Bincoletto G, 'The EDPB-EDPS Joint Opinion on the Commission Proposal for a Regulation on the European Health Data Space: Key Issues to Be Considered in the Legislative Process' (2022) 8 Eur. Data Prot. L. Rev. 398.

Botrugno C and others, 'Cybersecurity, privacy and health data protection in the digital strategy of the European Union' (2022) 3 RECHTD. Revista De Estudos Constitucionais, Hermenêutica E Teoria Do Direito 300.

Chiappetta A, 'Proposta di studio clinico sugli esiti della colecistectomia nel Veneto' (Presentation at the Best specialization dissertation Prize event, Società Triveneta di Chirurgia, 21/12/2018) – courtesy of the author.

Cohen IG and others, 'The European artificial intelligence strategy: implications and challenges for digital health' (2020) 2(7) The Lancet Digital Health e376.

Crighton E, 'Public health screening programme annual report: 1 April 2018 to 31 March 2019' [2020].

Custers B and Malgieri G, 'Priceless data: why the EU fundamental right to data protection is at odds with trade in personal data' (2022) 45 Computer Law & Security Review 105683.

Dantas C and Mackiewicz K, 'Are we ensuring a citizen empowerment approach for health data sharing?', in Anto Čartolotovni and others (eds.) *Proceedings of the 2022 GoodBrother International Conference on Privacy-friendly and Trustworthy Technology for Society* (2022) 55.

DeFrancesco L and Klevecz A, 'Your DNA broker' (2019) 37(10) Nat Biotechnol 842.

Dhirani LL and others, 'Ethical dilemmas and privacy issues in emerging technologies: a review' (2023) 23(3) Sensors 1151.

EDPB E, 'EDPB-EDPS Joint Opinion 2/2022 on the Proposal of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)' [2022] EDPB/EDPS, Brussels.

Elmlinger K, 'My Data, My Choice? An Analysis of the Data Broker Industry and the Exercise of Privacy Choices' [2023] Princeton University Undergraduate Senior Theses.

Eruchalu CN and others, 'The expanding digital divide: digital health access inequities during the COVID-19 pandemic in New York City' (2021) 98 Journal of Urban Health 183.

Ferretti F, 'Directive (EU) 2019/770: personal data as consideration in contracts for the supply of digital content and digital services and the inherent impact on privacy law' (2021) Actualidad Jurídica Iberoamericana 16, 1740–1777.

Finck M and Pallas F 'They who must not be identified – distinguishing personal from non-personal data under the GDPR' (2020) 10(1) International Data Privacy Law 11.

Gallego GF, De Ampuero Castellanos S and Robles JR, 'TSending personal data, receiving non-personal data: Recent EU judgment reinforces the power of pseudonymization' [2023] Lexology.

Gallese C, "Legal Issues of the Use of Chatbot Apps for Mental Health Support" (2022).

Gallese C, "Predictive justice in light of the new AI Act proposal" [2022] available at SSRN <https://ssrn.com/abstract=4286023> accessed 21 May 2024.

Gallese C, "A first commentary to the proposal for a new Regulation on fair access and use of data (Data Act)" (2022) 3 Media Laws.

Gonzalo F G and others, 'Sending personal data, receiving non-personal data: Recent EU judgment reinforces the power of pseudonymization' [2023] Lexology

Gordon WJ and others, 'Beyond validation: getting health apps into clinical practice' (2020) 3(1) NPJ digital medicine 14.

Graham M, 'Data for sale: trust, confidence and sharing health data with commercial companies' (2023) 49(7) Journal of Medical Ethics 515.

Grundy Q, 'A review of the quality and impact of mobile health apps' (2022) 43 Annual review of public health 117.

Hughson J-aP and others, 'The rise of pregnancy apps and the implications for culturally and linguistically diverse women: narrative review' (2018) 6(11) JMIR mHealth and uHealth e189.

International Privacy, 'How Apps on Android Share Data with Facebook (even if you don't have a Facebook account)' (Privacy International 2018).

Kazmierczak M and others, *China's Biotechnology Development: The Role of US and Other Foreign Engagement* (US-China Economic and Security Review Commission 2019).

Kim J, 'Data Brokers and the Sale of Americans' Mental Health Data' (Duke University 2022).

ICO, "Generative AI third call for evidence: accuracy of training data and model outputs", ICO consultation series on generative AI, available at <https://ico.org.uk/about-the-ico/what-we-do/our-work-on-artificial-intelligence/generative-ai-third-call-for-evidence/> accessed 21 May 2024.

Lee Y and Moon M, 'Utilization and content evaluation of mobile applications for pregnancy, birth, and child care' (2016) 22(2) Healthcare informatics research 73.

Li W and Quinn P, 'The European Health Data Space: An expanded right to data portability?' (2023) 52 Computer Law & Security Review.

Margarida M and others, 'Implementation Status of the Proposal for a Regulation of the European Health Data Space in Portugal: Are We Ready for It?' (2023) 302 Studies in Health Technology and Informatics 48.

Marelli L and others, 'The European health data space: Too big to succeed?' (2023) 135 Health policy 104861.

Martínez-Pérez B, De La Torre-Díez I and López-Coronado M, 'Privacy and security in mobile health apps: a review and recommendations' (2015) 39 Journal of medical systems 1.

Mitchell UA and others, 'The digital divide in health-related technology use: The significance of race/ethnicity' (2019) 59(1) The Gerontologist 6.

Mursia M and Trovato CA, 'The commodification of our digital identity: limits on monetizing personal data in the European context' (2021) 2 Media Laws 165.

Narayanan A and Shmatikov V, 'Myths and Fallacies of Personally Identifiable Information' (2010) 53 Communications of the ACM 24, 26.

Narayanan A and Shmatikov V, *Robust De-anonymization of Large Sparse Datasets* in Proceedings of the 2008 IEEE Symposium on Security and Privacy (IEEE Computer Society 2008) 111.

Ohm P, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCL Law Review 1701.

Petrocnik T, 'Health data between improving health (care) and fuelling the data economy' (2022) 2022 Technology and Regulation 124.

Pinchot J and others, 'Data Privacy Issues in The Age of Data Brokerage: an Exploratory Literature Review' (2018) 19(3) Issues in Information Systems.

Poletti D, 'Le condizioni di liceità del trattamento dei dati personali' [2019] Giurisprudenza italiana 2783.

Privacy International, '*How digital health apps can exploit users' data'* (Privacy International 2022).

Public Interest Advocacy Centre, Mental health discrimination in insurance (Public Interest Advocacy Centre, Liverpool, Sydney 2021).

Ricciuto V and others, 'La patrimonializzazione dei dati personali. Contratto e mercato nella ricostruzione del fenomeno' in *I dati personali nel diritto europeo* (Torino 2019).

Riso B and others, 'Ethical sharing of health data in online platforms–which values should be considered?' (2017) 13 Life sciences, society and policy 1.

Rocher L, Hendrickx JM and De Montjoye YA, 'Estimating the success of reidentifications in incomplete datasets using generative models' (2019) 10(1) Nature communications 1.

Rössler B and others, 'Should personal data be a tradable good? On the moral limits of markets in privacy' [2015] Social dimensions of privacy: Interdisciplinary perspectives 141.

Senigaglia R, 'La dimensione patrimoniale del diritto alla protezione dei dati personali' (2020) 2(2) Contratto e impresa 760.
Shabani M, 'Will the European Health Data Space change data sharing rules?' (2022) 375(6587) Science 1357.
Silbey SS, 'After legal consciousness' (2005) 1 Annu. Rev. Law Soc. Sci. 323.
Sweeney L, 'Weaving technology and policy together to maintain confidentiality' (1997) 25(2–3) The Journal of Law, Medicine & Ethics 98.
Thomason J, 'Big tech, big data and the new world of digital health' (2021) 5(4) Global Health Journal 165.
Thornicroft G and others, 'The Lancet Commission on ending stigma and discrimination in mental health' (2022) 400(10361) The Lancet 1438.
Veale M and others, "When Data Protection by Design and Data Subject Rights Clash" (2018) 8 International Data Privacy Law 105, 107.
Weitzenboeck EM and others, 'The GDPR and unstructured data: is anonymization possible?'(2022) 12(3) International Data Privacy Law 184.
Yakovleva S and Irion K, 'Toward Compatibility of the EU Trade Policy with the General Data Protection Regulation' (2020) 114 American Journal of International Law 10.