EUROPEAN JOURNAL OF HEALTH LAW 30 (2023) 345–363

BRILL | NIJHOFF

European Journal of Health Law

brill.com/ejhl

Compromises and Asymmetries in the European Health Data Space

Petros Terzis | ORCID: 0000-0001-8985-5651
Faculty of Law, University College London, Bentham House, Office 335, London WC1H 0EG, UK
petros.terzis@ucl.ac.uk

Abstract

In the post-pandemic world, the ability of researchers to reuse, for the purposes of scientific research, data that had been collected by others and for different purposes has rightfully become a policy priority. At the same time, new technologies with tremendous capacity in data aggregation and computation open new horizons and possibilities for scientific research. It is in this context that the European Commission published in May 2022 its proposal for a sector-specific regulation aiming at establishing the legal landscape and governance mechanisms for the secondary use of health data within the European Union. The ambitious project is centred on administrative efficiency and aspires to unleash the potential of new technologies. However, the quest for efficiency usually comes with privacy compromises and power asymmetries and the case of the European Health Data Space Regulation is no different. This paper draws attention to some of these compromises and suggests specific amendments.

Keywords

Big Tech – European Health Data Space – health data – secondary use – technology

1 Introduction

The European Union is moving towards establishing the foundations for the common European Health Data Space. The project aspires to unleash the full potential of health data and as such it will become part of a much broader legal context comprised of predominantly horizontal legal and governance

Published with license by Koninklijke Brill NV | DOI:10.1163/15718093-BJA10099

frameworks: the General Data Protection Regulation (GDPR), the Data Governance Act (DGA), the (draft) Data Act, and the Network and Information System Directive.[1] Within this context, the European Commission published its proposal for the European Health Data Regulation (hereinafter 'Proposal'). The Proposal has 76 articles, and its main core can be divided into two pillars. The first pillar (Article 1–32) covers issues related to primary health data (most notably the design and development of electronic health registers, the interoperability of electronic health record systems across the EU Member States, and rules for wellness applications) whilst the second (Article 32–58) establishes for the first time the legal framework for the secondary use of health data within the EU. Although there are interesting aspects to discuss in the first pillar (particularly with regards to the self-regulatory powers for developers of wellness applications) this article is focusing on the second pillar. Through a systematic analysis of the provisions therein, it draws attention to the latent fallacy of equating scientific research, on the one hand, and algorithmic projects on the other, the problems that such consonance can generate, as well as the need for data subjects to be meaningfully informed about the fate of their health data. The note then moves on to recommend amendments in accordance with the established rules and principles of data protection as balanced by the need to promote scientific research and innovation.

2 Background

Health data, as a sensitive category of data, enjoys a high threshold of legal protection. At the same time, the secondary use of such data, meaning the use of health data that have been collected for primary use under different justifications and legal bases, can offer benefits for healthcare provision and research. For example, aggregated and anonymised health data that have been collected from electronic health records or clinical trials and under different legal bases (ie explicit consent or public interest) can be reused to — amongst others — promote scientific research, inform evidence-based policymaking,

1 Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) 2016; Regulation 2022/868 of the European Parliament and of the Council on European Data Governance (Data Governance Act) 2022; Proposal 2022/0047 (COD) for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) 2022; Directive (EU) 2016/1148 of the European Parliament and of the Council concerning measures for a high common level of security of network and information systems across the Union 2016.

or improve the efficiency of healthcare systems across populations through better allocation of resources. Likewise, sharing health data across institutions and databases can validate whether research findings are institution-specific or can lead to generalisable outcomes.[2]

In such a context, the COVID-19 pandemic fuelled the debate on the need for cross-border cooperation in public health emergencies with health data being at the epicentre of it. Scholars have argued that the lack of a common framework for the use and re-use of electronic health data have posed key barriers to COVID-19 scientific research.[3] Therefore, the secondary use of health data has rightfully become a policy priority. As Recital 38 of the Proposal acknowledges:

> [M]uch of the existing health-related data is not made available for purposes other than that for which they were collected [...] In order to fully unleash the benefits of the secondary use of electronic health data, all data holders should contribute to this effort in making different categories of electronic health data they are holding available for secondary use.

The secondary use of personal data, in general, is currently covered by GDPR Article 6 (4) and the Article 29 Working Party Guidelines on purpose limitation.[4] For health data, in particular, governments across the EU and around the world have established data infrastructures and — less often — legal frameworks for enabling its secondary use.[5] In this direction, the Proposal by the European Commission attempts to create a data governance regime that will offer a less costly alternative to consent as a legal basis for the collection and processing of electronic health data while, in the meantime, paving the way for cross-border interoperability of health data infrastructures. This new regime is expected to promote the ability of researchers, policymakers, and doctors across the EU to

2 S. McLennan, S. Rachut, J. Lange, A. Fiske, D. Heckmann and A. Buyx, 'Practices and Attitudes of Bavarian Stakeholders Regarding the Secondary Use of Health Data for Research Purposes During the COVID-19 Pandemic: Qualitative Interview Study', *Journal of Medical Internet Research* 24 (2022) e38754, citing L.A. Celi and others, '"Big Data" in the Intensive Care Unit. Closing the Data Loop', *American Journal of Respiratory and Critical Care Medicine* 187 (2013) 1157–1160.

3 McLennan et al., *supra* note 2; S. McLennan, L.A. Celi and A. Buyx, 'COVID-19: Putting the General Data Protection Regulation to the Test', *Journal of Medical Internet Research Public Health and Surveillance* 6 (2020) e19279.

4 Article 29 Working Party, Opinion 03/2013 on purpose limitation, pp. 23–28.

5 For an overview of the secondary use of health data in Europe see M. Boyd, J. Tennison and A. Alassow, *Secondary Use of Health Data in Europe* (London: Open Data Institute, 2021), available online at https://theodi.org/wp-content/uploads/2021/09/Secondary-use-of-Health-Data-In-Europe-ODI-Roche-Report-2021-5.pdf.

use and reuse health data for different purposes, including research, innovation, policymaking, patient safety or personalised medicine.[6]

3 The Proposal

In this spirit, the Proposal is a first attempt to solidify a common EU framework for the secondary use of health data. As such, it could be viewed as a specialised framework for data governance within the broader landscape established by the DGA. Provisions between the two regimes are often overlapping as it is the case, for example, with the DGA's single information points and the Proposal's introduction of the national datasets for electronic health data.[7] Likewise, the Proposal introduces the concept of data altruism in health with explicit references to the respective provisions of the DGA.[8] More importantly, however, where the DGA introduces a generic framework for secondary use of public sector data, the Proposal creates a legal right for such use in the domain of electronic health data.[9] Likewise, where the (draft) Data Act provides for the use by the public sector of data held by private entities only in cases of public emergencies, the Proposal moves a small step beyond that to allow public sector bodies to obtain access to information that they require for fulfilling their tasks assigned to them by law.[10]

In terms of definitions, by 'secondary use', the Proposal refers to the processing of electronic health data which 'may include personal electronic health data initially collected in the context of primary use, but also electronic health data collected for the purpose of the secondary use'.[11] More importantly, by 'health data' the Proposal covers a broad range of fifteen categories of electronic health data enlisted in Article 33. These are categories of data that are not only strictly linked (ie electronic health records or genomic data), but also remotely related to health and care (ie electronic data related to insurance status, professional status, education, lifestyle, wellness and behaviour). Recital 39 explains this broad approach by indicating that: '[t]he categories of electronic health data that can be processed for secondary use should be broad and flexible enough to accommodate the evolving needs of data users, while

6 Proposal 2022/0140 (COD) for a Regulation of the European Parliament and the Council on the European Health Data Space pt Recital 38.

7 *Ibid*., Article 37.1 (q) (i).

8 *Ibid*., Recital 45, Article 40.

9 *Ibid*., Article 4.

10 *Ibid*., Article 48.

11 *Ibid*., Article 2 (2)(e).

remaining limited to data related to health or known to influence health'.[12] The recital then moves on to provide examples of health data that may be valuable on secondary use, albeit non-directly related to health or care. This may include '[…] consumption of different substances, homelessness, health insurance, minimum income, professional status, behaviour […]'.[13]

To understand the policy choice of maintaining such a broad categorisation of electronic health data, we need to view it within the context of the broader EU strategy for digital transformation and the plan for a European Digital Identity Framework and Wallet which will enable citizens to have trustworthy cross-border access to their health data from mobile devices.[14] As part of this project, the Proposal aims at enlarging the common data resources by seeking to pool and connect not only administrative data, data from medical devices, or genomic data, but also electronic data from digital health applications as well as data related to professional status, employment status, and lifestyle.

For the aggregation and management of access to electronic health data for secondary use, the Proposal introduces a novel data governance mechanism with broad powers and responsibilities, the health data access bodies. Each Member State will have either one or more such bodies. These will be independent bodies (subject to financial monitoring and judicial review) funded by Member States while the costs of their operations will be partially offset through fees charged for data applications and usage.[15] Health data access bodies will be explicitly encouraged to cooperate with supervisory authorities and stakeholders' representatives (ie patient organisations, health professionals, and researchers).[16] There seems to be neither an explicit reference that staff members of the health data access bodies will be civil servants, nor a limitation on who can participate in the health data access bodies. Generally, though, 'staff of the health data access bodies shall avoid any conflicts of interest' while Article 36 (6) indicates that '[h]ealth data access bodies shall not be bound by any instructions, when making their decisions'. As explained below, these bodies will essentially be in a position to collect information and centralise

12 As explained below, by 'data users' the proposal does not refer to 'data subjects' but to natural or legal persons that, following the issuance of a data permit, are allowed to access health data from the health data access bodies.

13 *supra* note 6, Recital 39.

14 European Commission, *Communication from the Commission — A European Health Data Space: Harnessing the Power of Health Data for People, Patients and Innovation* (Brussels: European commission, 2022), available online at https://ec.europa.eu/health/publications/communication-commission-european-health-data-space-harnessing-power-health-data-people-patients-and_en (accessed 6 June 2022).

15 *Supra* note 6, p 12.

16 *Ibid*., Article 36 (1).

information about electronic health databases from a plethora of resources and applications, public and private. Following the aggregation of information on the available databases, they will mediate and moderate access to health data according to the rules set out in the Proposal. According to Recital 55 of the Proposal, the processing of health data by data users is expected take place within a secure processing environment that will have the technical safeguards necessary to reduce privacy risks and 'prevent the electronic health data from being transmitted directly to the data users'. Finally, the Proposal also includes provisions for the secure transfer and processing of health data to and from the health data access bodies' databases.[17]

The health data access bodies will enjoy a range of duties, powers, and responsibilities serving essentiality as *de facto* health data managers and administrators. In particular, they will be responsible for: deciding on health data applications, preserve the confidentiality of IP rights, manage the infrastructure where health data is stored and processed, make information related to their databases accessible to the public, supervise data holders (entities which hold health data) and data users (entities which are given access to health data) and impose penalties pursuant to the provisions of the proposal.[18]

In theory, the process for the management of health data for secondary use within Member States will proceed as follows: Any entity that offers services or performs research in the health or care sector and happens to hold data which fall into one or more of the 15 categories enlisted in Article 33 will be obliged to provide information about this data to the designated health data access body (or bodies) of a Member State. It is up to the Commission to define through implementing acts 'the minimum information elements data holders are to provide for datasets and their characteristics'.[19] The health data access body will then compile and make public 'a national dataset catalogue that shall include details about the source and nature of electronic health data' as well as 'the conditions for making electronic health data available'.[20] Following that, any natural or legal person will be able to submit an application to the health data access body in order to be granted a 'data permit' according to the provisions and purposes of the Proposal. The health data access body will have 30 days to respond to data applications and, '[w]here a health data access body fails to provide a decision within the time limit, the data permit shall be

17 *Ibid.*, Article 50.
18 *Ibid.*, Article 37.
19 *Ibid.*, Article 55.
20 *Ibid.*, Article 37 (1)(q)(i).

issued'.[21] Once granted, the health data access body will coordinate access to this data through a secure processing environment without data leaving that depository. The data user and the health data access body will be joint controllers of the data made available under a particular data permit.[22] The accessed data will be anonymised unless the applicant provides explanations on why access to the data is required in a pseudonymised format.[23] Failure to cooperate in good faith with the health data access bodies could lead to a fine or ban from participating in the European Health Data Space. Finally, the access period is set for up to five years at the end of which the data users will be able to either extent their access for a maximum of another five years or, following guidance by a health data access body, 'store the dataset in storage system with reduced capabilities' in order to reduce access costs and fees.[24]

Article 46 (11) provides that the entities which use health data shall make public 'the results or output of the secondary use of electronic health data [...] no later than 18 months after the completion of the electronic health data processing', while Article 35 offers some safeguards for data subjects by prohibiting the accessing and processing of health data for secondary use for taking decisions detrimental to a natural person (including decisions on insurance or other benefits) or for advertising and marketing activities.

In terms of the legal basis, the proposed regulation explicitly acknowledges that for the exchange between data holder and the health data access body the legal basis is GDPR Article 6 (1) point (c) and Article 9 (2)(h), (i), and (j). In turn, the Proposal recognises that the legal basis for requesting access to the health data will be GDPR Article 6 (1)(e) and (f).[25] This essentially means that a data applicant will either rely upon the legal basis of a task carried out in the public interest, in which case the data application shall make reference to another EU or national law mandating the applicant to process health data for the compliance of its tasks; or it will rely upon the applicant's legitimate interests. In the latter case, the Proposal itself will serve as the guarantor of such claim and the decision of the health data access body, as the Recital explicitly acknowledge, will merely be an administrative decision determining the conditions for access to the requested data.[26]

21 Proposal 2022/0140 (COD) for a Regulation of the European Parliament and the Council on the European Health Data Space (n 6) s 46(3).

22 *Ibid.*, Article 51.

23 *Ibid.*, Articles 44 (3), 45 (2) (d), 45 (4), 45 (5).

24 *Ibid.*, Article 46 (9).

25 *Ibid.*, Recital 37.

26 *Ibid.*

Overall, the health data access bodies represent the EU's policy towards a model of health data governance with standardised infrastructures across its Member States along with a framework for their cross-border interoperability. A new governance model is thereby created. In this new space, decentralised health data access bodies will control and manage access to rather broad categories of health data generated by various actors within a Member State whilst all data will be subsequently integrated into the HealthData@EU, the cross-border infrastructure for secondary use of electronic health data. These interactions as well as the development of the required infrastructure will be facilitated and monitored by a newly established European Union body, the European Health Data Space Board which will initiate and supervise the task of coordinating the design of standards for interoperability of health databases across Member States.

4 Asymmetries and Compromises

Looking back to the short history of regulating data-intensive sectors and domains, one pattern that emerges is that, usually, the quest for efficiency and optimisation comes with privacy compromises and power bargains. From workplaces to national borders, and from ID verification to loyalty programs, surveillance technologies and information systems have been deployed to capture and analyse human behaviour to inform policy and optimise services.[27] The Proposal can thus be read in a similar framework as it attempts to strike a balance between the benefits of secondary use of health data on the one hand and the protection of data subjects' fundamental rights on the other.

Generally, storage and transfer of sensitive data across databases increases risks for data breaches, whilst the more remote data becomes from its source the less likely it is for data subjects to be aware of who accessed what data and for what purposes. In parallel, information deriving from health data may, once identified, lead to decisions or actions that are detrimental to individuals (ie denial or increase of insurance premiums, credit and mortgage status,

27 See, indicatively, R.A. Bales and K.V.W. Stone, 'The Invisible Web at Work: Artificial Intelligence and Electronic Surveillance in the Workplace', *Berkeley Journal of Employment and Labor Law* 41 (2020) 1–60; P. Molnar, 'Territorial and Digital Borders and Migrant Vulnerability Under a Pandemic Crisis', in: A. Triandafyllidou (ed.), *Migration and Pandemics: Spaces of Solidarity and Spaces of Exception* (Cham: Springer International, 2022) pp. 45–64; S. Milan, M. Veale, L. Taylor and S. Gürses, 'Promises Made to Be Broken: Performance and Performativity in Digital Vaccine and Immunity Certification', *European Journal of Risk Regulation* 12 (2021) 382–392.

direct marketing of medical products etc). At the same time, some health data is valuable if they allow some form of identification (ie data related to disease patterns). Anonymisation and aggregation may only partially prevent some of these challenges as the risk of re-identification is omnipresent and particularly acute in cases where health data is collected and processed by Big Tech companies with immense data wealth.

Analysing aspects of the Proposal in light of these general risks is important but remains out of the scope of this paper. Instead, this paper focuses on the Proposal's hastily designed provisions and controversial policy choices reflected on the Commission's active — and to a certain degree unjustified — commitment to technology companies and their AI/Algorithmic prospects and projects; a commitment once reserved exclusively for (public and private) scientific research.

4.1 *Health Data: By Whom and For Whom?*

Not all actors in the European Health Data Space are bearers of the same rights and obligations. There is a discrepancy between who is a data holder and who can become data user. More specifically, in narrowing down the remit of the entities which will bear an obligation to share data following the issuance of a data permit, Article 33 (3) of the Proposal indicates:

> The electronic health data referred to in paragraph 1 shall cover data [...] collected by entities and bodies in the health or care sectors, including public and private providers of health or care, entities or bodies performing research in relation to these sectors, and Union institutions, bodies, offices and agencies.

In the same spirit, 'data holder' is defined by the Proposal as 'any natural or legal person, which is an entity or a body in the health or care sector, or performing research in relation to these sectors[...]'.[28] The European Data Protection Board and European Data Protection Supervisor Joint Opinion on the Proposal (hereinafter 'Joint Opinion') warns that such phrasing may create inconsistencies and confusion with DGA's more generic definition of 'data holder'.[29] More

28 *Supra* note 6, Article 2 (2)(y).

29 'EDPB-EDPS Joint Opinion on the Proposal for a Regulation on the European Health Data Space' (European Data Protection Board) para. 44. Article 2 para. 8 of the DGA defines 'data holder' as 'a legal person, including public sector bodies and international organisations, or a natural person who is not a data subject with respect to the specific data in question, which, in accordance with applicable Union or national law, has the right to grant access to or to share certain personal data or non-personal data'.

importantly, however, given the Proposal's definition, it is unclear whether digital infrastructures with immense power and control over databases valuable for scientific research in healthcare will indeed be required to share data with interested parties. Are Google or its subsidiary, DeepMind, private entities performing research in relation to health or care sector or are they to be regarded as 'technology companies' and 'AI companies', respectively? Could Facebook's 'Reality Lab', Microsoft's 'Health Futures', or Amazon's 'AWS for Health' be encompassed by these provisions? Is Apple an entity performing research in health or care sector? Could we regard Apple's electrocardiogram (ECG) application for the iWatch as a healthcare service for that purpose? What about the plethora of health applications (such as meditation or period tracking applications) that are not 'entities performing research' per se but play a major role in amassing health data that seem to fall squarely under the Proposal's categorisation and could potentially support scientific research, statistical purposes, and evidence-based policymaking? Will (any of) these entities be obliged, following a data request, to share the data they collect with the health data access bodies and subsequently with the data user?

One could rightfully argue that, in principle, narrowing down the nature of the entities that will be obliged to share health data for secondary use seems logical. What seems to be at odds with these provisions, however, is the nature of the entities at the other side of the spectrum. For contrary to the criterion of relevance to health or care for the obligation of making data available to the health data access body, the Proposal does not restrict the remit of the natural or legal persons that can apply for access to health data. Instead, it allows 'any natural or legal person' to submit a data request provided that the applicant fulfills one or more of the purposes enlisted in the Proposal (see below).[30] As a result, the Proposal adopts a rather open-ended classification for the entities that will be entitled to access health data, whereas at the same time it contains the range of the 'data holders' to those entities that belong in the sphere of healthcare (either as provider of such services or as researchers). This raises the following questions: What is the policy rationale behind this double standard? Is there an explicit expectation from entities outside of the healthcare sphere to contribute to health research and innovation? If so, why not treat such entities as 'data holders' as well and not only potential 'data users'?

4.2 *Health Data for What?*

The problem of expanding the pool of potential data users beyond health or care becomes even larger when considering the Proposal's provisions about

30 *Supra* note 6, Article 47.

the purposes that can support a data application for access to electronic health data for secondary use. The list of purposes is found in Article 34. There, following provisions 34 (1)(a) to (e) which seem to be in line with the existing GDPR provisions for the collection and processing of sensitive data, we read that health data access bodies shall provide access to electronic health data where the intended purpose of processing complies with:

> (f) development and innovation activities for products or services contributing to public health or social security, or ensuring high levels of quality and safety of health care, of medicinal products or of medical devices;
>
> (g) training, testing and evaluating of algorithms, including in medical devices, AI systems and digital health applications, contributing to the public health or social security, or ensuring high levels of quality and safety of health care, of medicinal products or of medical devices;
>
> (h) providing personalised healthcare consisting in assessing, maintaining or restoring the state of health of natural persons, based on the health data of other natural persons.

We can read points (f) and (g) as encompassing anything that could fit under the already over-stretched umbrella of 'artificial intelligence' while point (h) as including (or directly referring to) the data-intensive industry of healthcare Internet of Things (wearables, sensors, actuators etc). Articles 34 (1)(f–h) should be read along with Article 37 (1)(i) which enlist the support of 'AI systems, the training, testing and validating of AI systems and the development of harmonised standards and guidelines [...] for the training, testing and validation of AI systems in health' as explicit obligations of the health data access bodies. In parallel, Article 33 of the Proposal which enlists the various categories of electronic health data, exacerbates the problematic character of Article 34 para. 1 (f–h). This is because an already overstretched categorisation of electronic health data will meet an open-ended basis for access to such data. Under the proposed regime, for example, it would not be difficult to imagine a tech company submitting — and perhaps — being granted a data permit for accessing data from insurance companies as well as education and meditation applications in order to develop a personalised recommendation system for 'healthy lifestyle'. The 'Joint Opinion' is thus right in recommending: a) the need for compatibility of Article 34 (1)(f–g) of the proposal with GDPR Article 9 (2); and b) the striking out categories of data enlisted in Article 33 (1)

and in particular points (f) and (n) which cover person-generated electronic health data and data related to insurance status, professional status, and other data relevant to wellness and behaviour, respectively.[31]

It is important to view Article 34 in this systematic way. Because it is in such a context that Article 34 in combination with Article 37 and Article 33 of the Proposal create a latent backdoor for the accessing of electronic health data by Big Tech or by any other entity with the infrastructural, logistical, and financial capacity to experiment with machine learning, algorithms, and personalised 'smart' technologies. Under the proposed regime, any natural or legal entity with the technical and infrastructural capacity to support a data application for 'training, testing, and evaluating of algorithms' or 'personalised healthcare' will have the option of getting access to, and processing troves of broadly defined health data without having to meet any of requirements set out in GDPR Article 9 (2). Instead, access to health data will be possible through an administrative pathway initiated by a data application that will — amongst other declaratory remarks — include:

> (e) a description of the safeguards planned to prevent any other use of the electronic health data;
>
> (f) a description of the safeguards planned to protect the rights and interests of the data holder and of the natural persons concerned;[32]

The Proposal thus generates concerns of procedural justice and changes the landscape upon which legal and policy dialogue takes place. In particular, by allowing secondary use of health data for 'training, testing and evaluating of algorithms [...] in [...] AI systems and digital health applications', Big Tech is offered new ways and bases for getting access to health data and for leveraging their expertise to the, non-native to theirs, health domain.[33] For example, under the new Proposal, a technology company will neither have to undergo the burdensome process of requesting the explicit consent of the data subjects whose electronic health data it wishes to use for its research activities, nor will it need to build a wholly speculative case around potential 'research exemption' for a particular project. Rather, it will be able to access data in a lawful manner and to evade public scrutiny by merely pleading its case about

31 *Supra* note 29, paras 36 and 90.

32 *Supra* note 6, Article 45 (2)(f).

33 T. Sharon, 'When Digital Health Meets Digital Capitalism, How Many Common Goods Are at Stake?', *Big Data & Society* 5 (2018), DOI: 10.1177/2053951718819032; T. Sharon, 'From Hostile Worlds to Multiple Spheres: Towards a Normative Pragmatics of Justice for the Googlization of Health', *Medicine, Health Care and Philosophy* 24 (2021) 315–327.

its algorithmic project and the measures it will put in place to protect people's data. Similarly, given the administrative nature of the procedure for accessing health data and the broad approach in defining it, data applications of unclear scientific validity or even pseudoscientific research endeavours (such as 'Emotional AI' applications) will no longer be judged on scientific merit as long as they will be capable of becoming part of an application that fits the criteria set out in the Proposal.

This latent backdoor engenders problems, risks, and challenges for several other reasons. Clinical trials and medical research aim at assuring the generality of their results. Algorithmic projects cannot replicate this quest for generality through pilot studies and limited datasets. This inherent characteristic of machine learning may be viewed under the new regime as a legitimate justification for multiple data access applications and extensive use of health data on behalf of Big Tech. In parallel, the provisions about the access period (5 years further extendable for up to 5 years) coupled with the AI systems' continuous need for 'data fuel' as well as their organic tendency to search for problems rather than solutions of particularised nature, may ultimately allow Big Tech companies to justify their need for the maximum access period allowed by the Proposal (10 years).[34] This may contradict the European Court of Human Rights jurisprudence on the need for heightened protection required for the retention of genetic and biometric data.[35] Besides, access to health data for secondary use by a technology company for a 10-year period cannot be based on the same policy and epistemic grounds as the wholly legitimate need for maintaining higher retention periods for medical records and clinical trial data.[36]

Finally, it is entirely unclear whether access to novel sources of data really responds to the challenges that machine learning and artificial intelligence in healthcare are supposed to confront. This is because, contrary to other areas of algorithmic decision-making systems where accurate interpretability may not be organically significant, physicians are likely to be far more interested in the thought process behind an outcome rather than the outcome itself.[37]

34 It is also entirely unclear what will be the fate of the used data after the end of the retention period.

35 *S and Marper v the United Kingdom* [2008] ECtHR [GC] 30562/04, 30566/04; *Amann v Switzerland* [2000] ECtHR [GC] 27798/95.

36 In these cases, it is usually patient continuity and possibility of significant late effects, respectively, that justify longer retention periods.

37 M.A. Ahmad, C. Eckert and A. Teredesai, 'Interpretable Machine Learning in Healthcare', *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (New York, NY: Association for Computing Machinery, 2018) pp. 559–560; H. Habibzadeh, 'A Survey of Healthcare Internet of Things (HIoT): A Clinical Perspective', *IEEE Internet of Things Journal* 7 (2020) 53–71, p. 62.

Providing access to more data does not solve this problem. If anything, it complicates it further.

In parallel, aside from the teleological argument there is also a normative one as the Proposal deviates from the established rules for the use of sensitive data (which include health data). This is because GDPR Article 9 (1)(h) to (j) carve out specific exceptions to the general prohibition on the collection and processing of sensitive data (such as health data). The clearly identified exceptions therein include among others: reasons of substantial public interest, preventive and occupational medicine, management of health or social care systems and services, cross-border public health threats, scientific and historical research, and other statistical purposes. In this context, it is generally accepted that the 'research exemption' of GDPR Article 9 (2)(h) to (j) and Article 89 already offers a broad and permissive regime for sensitive data collection and processing.[38] Scholarly work has emphasised on the need to bring clarity to the provision with some scholars arguing about strengthening the link between scientific research and the public interest more broadly; a link which the GDPR does not explicitly articulate.[39] Others developed frameworks for conceptualising and breaking down the content of the exemption.[40]

Asked to opine at the consultation period that preceded the publication of the Proposal, the European Data Protection Supervisor observed that any entity seeking access to health data through the proposed scheme shall 'be required to demonstrate specific objectives with scientific and research relevance with evident purposes of public interest […]'.[41] At no point does the EDPS make any reference to private entities experimenting with the use of algorithms in public health or private providers of personalised healthcare services. In the same spirit, the only country that has regulated the secondary use of health data is Finland whose 'Act on the Secondary Use of Health and Social Data' makes explicit reference to 'development and innovation activities' as legitimate grounds for secondary use by adding that a data permit on such grounds

38 European Data Protection Supervisor, *A Preliminary Opinion on Data Protection and Scientific Research* (Brussels: European Data Protection Supervisor, 2020); L. Marelli, G. Testa and I. Van Hoyweghen, 'Big Tech Platforms in Health Research: Re-Purposing Big Data Governance in Light of the General Data Protection Regulation's Research Exemption', *Big Data & Society* 8 (2021), DOI: 10.1177/20539517211018783.

39 *Supra* note 1, Recitals 157 and 159.

40 L. Floridi, 'Key Ethical Challenges in the European Medical Information Framework', *Minds and Machines* 29 (2019) 355–371.

41 European Data Protection Supervisor, *Preliminary Opinion 8/2020 on the European Health Data Space* (Brussels: European Data Protection Supervisor, 2020) para. 34.

must promote public health or social security, develop healthcare services, or protect the health and wellbeing of individuals.[42]

Fundamentally, despite its inherent ambiguity, what the 'research exemption' achieved was to focalise the point of enquiry primarily on epistemic grounds. That is, the decisive factor to determine the applicability of the 'research exemption' was whether a particular project could qualify as genuine science. It was then up to the data collectors to prove that they have the capacity and expertise to produce such work. This may have been easy and often self-evident for the case of pharmaceutical companies working on a new vaccine, but facing the same question, Big Tech and its experimental, data-intensive projects with AI systems and algorithms would struggle. For it is by no means self-evident that absent the Proposal's regime, Big Tech algorithmic projects involving health data would merit the scientific exemption. If anything, such projects were more likely to invite rigorous public scrutiny by journalists, academics, and other interested parties as was the case with the Deepmind's collaboration with Royal Free.[43] And although the 'research exemption' in GDPR set a rather broad benchmark with the recitals not helping much in bringing clarity, at the same time, it made clear to the entities seeking access to health data that the yardstick to be measured against is one's contribution to the advance of scientific knowledge.

None of the above, however, shall be regarded as a claim against the possibilities that 'artificial intelligence' (or simply advanced computation), or machine learning, or algorithms can offer for healthcare and medical science (and) research. Rather, it is a claim about holding the companies that want to build such technologies to the same epistemic and methodological standards with the rest of the scientific community that need access to health data for secondary use. The following question thereby arises: Why shall the law treat scientific research and algorithmic development as weighing equally in terms of their ability to justify an application for access to health data?

4.3 *Data Localisation and Consent*

Following the issuance of a data permit, health data access bodies will be responsible for providing access to health data through a secure processing environment.[44] Health data for secondary use will be processed within the developed repository — which may be managed by a third party — and data

42 Finnish Ministry of Social Affairs and Health, *Secondary Use of Health and Social Data* (Helsinki: Finnish Ministry of Social Affairs and Health, 2022), available online at https://stm.fi/en/secondary-use-of-health-and-social-data.

43 J. Powles and H. Hodson, 'Google DeepMind and Healthcare in an Age of Algorithms', *Health and Technology* 7 (2017) 351–367.

44 *Supra* note 6, Article 50.

users will only be allowed to process data within that environment and only download non-personal health data. However, as the 'Joint Opinion' warns, the fact that there is no explicit obligation for personal health data to be stored and processed within the EU may create fragmentations that could lead to different degrees of protection for data subjects across Member States.[45]

Finally, in contrast with GDPR Article 14, Article 38 (2) of the Proposal stipulates that health data access bodies will not be obliged to provide specific information to natural persons concerning the use of their data for particular projects under a data permit. Instead, they will provide general information, on a monthly basis, about all the data permits, requests, and applications they have received. The 'Joint Opinion' characterises Article 38 (2) of the Proposal an 'explicit derogation' from GDPR provisions on consent.[46] Although this deviation from the established regulatory framework may be justified by reasons of cost-efficiency, it is unclear why existing restrictions to the right of information laid out in GDPR Article 14 5 (b) and (c) (ie scientific research) would not be sufficient.[47] Following common practices in medical research management, the Proposal chooses an *ex-post* mechanism for providing general information by noting that health data access bodies will make public the data permit (or its response in any other case) within 30 working days of its issuance.[48]

As a result, the legal basis of informed consent, a requirement traditionally associated with the collection and processing of sensitive data, is thereby transformed into a transparency obligation of the newly established body. This is not by itself a bad idea as the complexity of medical research often justifies flexible arrangements at place for the use and reuse of health data. However, firstly, there are already the GDPR rules in place for such circumstances and, secondly, viewed in the context of the entities that may be granted a data permit, the respective provisions of the proposal merit extreme caution. This is because the wholly legitimate claim of a medical research group to be able to reuse health data collected as part of a clinical trial cannot be held on the same standard of transparency and flexibility with a technology company that requests access to data for the development of an AI system or for pursuing an experimental algorithmic project.

The problematic character of the provision becomes even more evident considering the history that technology companies have with engaging in

45 *Supra* note 29, paras 109–111.

46 *Supra* note 29, para. 23.

47 *Ibid.*, para. 25.

48 This is, for example, the case with applications submitted to the UK's NHS Health Research Authority. See, for example, the respective Confidentiality Advisory Group Registers (https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/confidentiality-advisory-group-registers/).

secret and controversial agreements with public sector bodies involving patient's data. As it stands, Article 38 (2) of the Proposal deprives interested parties (ie, data subjects, journalists, NGOs etc) from the ability to get informed in time and scrutinise imminent actions related to health data at a national or cross-border scale. The following questions thereby arises: Is Article 38 (2) compatible with the principles and spirit of the GDPR regarding transparency and information on processing of sensitive data? How is the Proposal's aim for supporting individuals to take control of their own data achieved and what measures have been put in place in this direction?

5 Conclusion and Recommendations

The Proposal brings about a seismic shift to the *status quo* of electronic health data within the EU. Boosted by the policy dynamics generated by the pandemic, it aspires — amongst others — to establish the policy and legal framework for the secondary use of health data in the EU. In doing so, it aims at unleashing the potential health data has for research, policymaking, and innovation. Regarding its negotiations stage, the Proposal is expected to reach the European Parliament later in the year and it is, currently, under discussion (1st reading) in the following committees: Civil Liberties, Justice, and Home Affairs Committee, Environment, Public Health, and Food Safety Committee, the Budget Committee (which has decided not to issue and opinion) and the Industry, Research, and Energy Committee.

Because of its extensive range, particular caution is required. For as it stands, the Proposal and the justifications for the secondary use of health data provided therein, disrupts the normative and regulatory roadmap for accessing and using electronic health data. The clearly articulated rules set out in GDPR are supplemented, and to a certain extent superseded, by a governance regime with different value-laden commitments and scope. Data protection of sensitive data is thus diluted into a governance system for administrative management of electronic health data. Within the latter, normative questions about access to health data are transformed into administrative questions to be dealt with purely compatibility criteria. This is not necessarily and by itself detrimental to people's fundamental rights. Societies can indeed benefit from the secondary use of health data. From scientific research to evidence-based policymaking, health data can contribute to the development of scientific knowledge, the rapid response to cross-border public health threats, or to educating the next generation of healthcare professionals.

But as we move towards the wholly legitimate goal of harnessing the power of electronic health data, we need to be aware of what is at stake and what

compromises are made. And as this paper has tried to articulate, the Commission's attempt to set the foundations of the European Health Data Space, has some rifts.

By creating an administrative backdoor for access to troves of health data that would not necessarily be justified under the GDPR (access for training algorithms or for development of personalised healthcare), the Proposal equates scientific research and other well justified reasons of public policy with experimental projects of different values and priorities whose primary scope is not the rigorous study of falsifiable hypotheses but rather a stray search in seek of opportunities. But equating the normative mandates of scientific research or cross-border public health threats with algorithmic experimentation in public health cannot be a mere result of administrative management or legal analysis.

For this reason, Article 34 (1) of the proposal and in particular points (f), (g), and (h) need to be substantially reviewed or entirely removed to be discussed at a later stage in the European Health Data Space lifecycle. For equating scientific research with projects of non-scientific priorities is, at its core, a political mistake that may risk the ultimate potential of an otherwise promising and much needed legal-political project. In the same spirit, the provisions allowing all natural and legal persons to submit a data request (and not only those entities in the health or care sectors) needs to be backed by a specific policy rationale or reviewed altogether especially in light of the antithetical provisions with regards to the nature of the entities that can be deemed as 'data holders'. Provisions and safeguards regarding anonymisation and pseudonymisation need to be considered carefully in light of the entities that may be given access to health data and their associated data wealth. Finally, making public a data permit (which essentially follows the disclosure of data to the applicant) within 30 days of its issuance does not bond well with the increased threshold of the required information and notification traditionally associated with collection and processing of sensitive data. An obligation from the part of the health data access body to make public the data application itself particularly in those cases where the data application is based on Article 34 (f) to (h) may remedy this anomaly.

The promise of AI and machine learning may indeed be significant for healthcare. But there is not enough scientific evidence to suggest that the progress of the existing private actors in the field of 'AI for healthcare' can indeed justify compromises of such scale. And even if there was some progress, such evidence shall be tested against the same standards with the rest of the scientific endeavours. Coupled with that, the AI and machine learning market and research field are heavily dependent on, and dominated by few technology companies whose data and computational infrastructures and logistics

enable data collection and processing as well as population-management functionality on a global scale.[49] Allowing such entities to access large pools of health data across the EU through a *quasi* self-assessment route that merely asks them to map their objectives to the scope of the Proposal, is likely to further consolidate and entrench their market power and political leverage for research and policy in the health domain. Against this current, a regulatory mindset, authentically committed to extend the boundaries of research possibilities for the European Health Data Space, could discuss the policy option of instituting a public mandate for companies with immense computational capacity (i.e., Google, Apple, Microsoft, and Amazon) to share their data and computing infrastructures for purposes of medical research.[50]

Whatever its current form may be, the Proposal offers an institutional opportunity for getting things right in building the legal framework of the promising European Health Data Space. But it also requires our vigilance. For, even if we assume that, as it stands, the Proposal is good for everyone, it is certainly much better for some.

Acknowledgement

Funding received from Fondation Botnar.

49 N. Ahmed and M. Wahed, 'The De-Democratization of AI: Deep Learning and the Compute Divide in Artificial Intelligence Research', *ArXiv* (2020), DOI: 10.48550/arXiv.2010.15581.

50 For an introduction to the problem of power engendered in digital infrastructures with advanced computational capacity, see S. Gürses and R. Dobbe, *Programmable Infrastructures* (Delft: TU Delft, 2021), available online at https://www.tudelft.nl/tbm/programmable-infrastructures (accessed 24 January 2022). For a discussion on the impact of private entities in shaping technologies for public health, see C. Troncoso, D. Bogdanov, E. Bugnion, S. Chatel, C. Cremers, S. Gürses, J.-P. Hubaux, D. Jackson, J.R. Larus, W. Lueks, R. Oliveira, M. Payer, B. Preneel, A. Pyrgelis, M. Salathé, T. Stadler and M. Veale, 'Lessons from a Pandemic: Deploying Decentralized, Privacy-Preserving Proximity Tracing', *Communications of the ACM* 65 (2022) 48–57.