PA PERSONA E AMMINISTRAZIONE
Ricerche Giuridiche sull'Amministrazione e l'Economia

DANIEL FOÀ
Ph.D. in Law and Business at LUISS Guido Carli
dfoa@luiss.it

SECONDARY USE OF HEALTH DATA AND ITS REGULATION

IMPACT ASSESSMENT OF THE PROPOSED EUROPEAN HEALTH DATA SPACE ON SCIENTIFIC RESEARCH

ABSTRACT

The proposed regulation (adopted by the EU Parliament) on the European Health Data Space provides for a simplified mechanism - to be managed by newly established bodies, responsible for access to health data and charged with a guarantee function - to make health data easily accessible and marketable for research purposes. This procedure would enable the sale of health data for secondary uses with scientific research purposes, without the direct involvement of the data subject. However, the data that can be shared must necessarily be anonymized or pseudonymised and their use must take place within secure processing environments.

The research questions the contribution aims to address is whether the model outlined in the regulation is compliant with the fundamental principles laid down in the GDPR – especially with regards to the rights of the data subject and the relevant safeguards – what consequences will this new paradigm produce on the market and if, after the entry into force of the regulation, there may still be any room for secondary use of fully personal health data.

KEYWORDS: Health data – EHDS – Secondary use – Scientific research – Pseudonymization – Anonymization

1. Introduction

1.1 The establishment of fully functional health data processing models is primarily aimed at ensuring the best medical care for patients to whom the data relate. It may also serve for further purposes related to the patient's own health (e.g., diagnostic or preventive purposes) and scientific progress (e.g., scientific research). Therefore, is essential that a clear legal framework exists as any interpretative uncertainty could harm the smoothness of the process: it needs to be efficient and not excessively burdensome (as it would easily discourage - at least - some of the possible use of the data) while granting the data subject sufficient control over the use of his health data.

The complex balancing between these instances has already been sought by the GDPR[1] and currently is under partial review in the context of the European Data Strategy[2].

In this regard, the proposal (now adopted by the EU Parliament) for a European Health Data Space (EHDS)[3] aims at establishing a model that -

---

1* A first version of the paper was presented at the international workshop "The Health of the Union" held at the University of Pisa on October 6, 2023. I would like to thank the participants of the seminar for their valuable comments. Since the very beginning of the debate, there have been varying opinions on the capacity of the GDPR regulation to facilitate and incentivise the use of health data for scientific research purposes. See, among the others, M.C. Ploem - M.L. Essink-Bot – K. Stronks, *Proposed EU data protection regulation is a threat to medical research,* BMJ 2013, 346: f3534.6; E.S. Dove - B. Thompson - B.M. Knoppers, *A step forward for data protection and biomedical research*, Lancet, 2016, 387:1374.

2 For a general introduction to the European Data Strategy and its implications see D. Amram, *Comparing EU initiatives on data: addressing risks and enhancing harmonisation opportunities*, Opinio Juris in Comparatione, n. 1/2023*;* PromethEUs, *The EU's Data Strategy from a multifaceted perspective. Views from Southern Europe,* June 2023; F. Bravo, *Intermediazione di dati personali e servizi di data sharing dal GDPR al Data Governance Act,* Contratto e impresa Europa 1/2021, p. 199 ff.

3 Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space, COM (2022) 197 final, 2022/0140(COD), May 3rd 2022. On March 15th 2024, a political agreement between the European Parliament and the Council of the EU has been reached on amended proposal and, on April 24th 2024, the EU Parliament has adopted it. On the topic see, among the others, D. Horgan et al., *European Health Data Space — An Opportunity Now to Grasp the Future of Data-Driven Healthcare*, *Healthcare* 2022, *10*, 1629, available at https://doi.org/ 10.3390/healthcare10091629; J.S. Marcus et al., *The European Health Data Space*, IPOL | Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament Policy Department studies, December 2022, available at https://ssrn.com/abstract=4300393; T. Schmitt et al., *What does it take to create a European Health Data Space? International commitments and national realities*, Z. Evid. Fortbild. Qual. Gesundh. wesen (ZEFQ), available at https://doi.org/10.1016/j.zefq.2023.03.011.

being aware of the potential of health data for multiple applications – pursues the goal of enhancing and fostering the quality of health data, promoting its use while strengthening the data subject trust: it is essential that every citizen feels confident that his data are correctly processed, in compliance with the rules on data protection, preserving confidentiality and security[4].

Such framework aims, firstly, to increase the quality and effectiveness of health care[5], allowing professionals to access health data relating to the patients, potentially achieving more accurate diagnosis and assessing the most appropriate therapeutic approach, even by simulating the efficacy of the treatment on the individual patient based on the available data and parameters[6]. The use of data gathered from clinical practice is essential to understand what the actual results (on a set of real patients) are in the administration of a therapy and management of diseases, what limitations and benefits exist, thereby shape the "best practices". Secondly, as hinted, health data may also serve many other purposes and among them scientific research has a prominent role.

1.2 The processing of health data may be subject to a range of regulations according to its classification: in turn, it may be a primary processing of data, when data are collected in the first instance for that specific purpose; or secondary processing, when data collected for different objectives

4 Health sector has its own peculiarities and motives that require a smoother mechanism for data use and re-use, but also imposes a heightened duty of confidentiality and security for data: the knowledge of information about the patient's health and particular medical conditions makes the patient particularly vulnerable to third parties.

5 The EU health commissioner Stella Kyriakides in June 2022 remarked that "*health data has vast potential to improve the health of our citizens if we manage to use it in an effective and safe way* [...] *sharing data will save lives. EHDS will help innovation in healthcare, which is all the more crucial as new health threats arise*". With reference to its potential relating to research she added that "*it will prove far easier to conduct research on a European scale with higher quality interoperable data*". See European Commission, *Opening Remarks by Commissioner Stella Kyriakides at the EPSCO Council—European Health Data Space,* 2022, available at https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_22_3708.

6 For example, building on the available data, it is possible to elaborate digital twins that enable physicians to simulate treatments of diseases in order to assess the relevant outcome on a specific individual.

(*e.g.* observational, clinical) are used for research[7]. Underlying the primary use of data for health care purposes are interests for its most agile and timely use, whilst ensuring the respect of the rights and interests of the data subject to whom the data relate; whereas the reasons justifying a specific regulation of the secondary processing of health data are different: incentivizing the provisions of new services and - most relevant from a public interest point of view - enabling and favouring health research[8]. Health data collected by healthcare institutions in the context, e.g., of diagnostic tests and healthcare treatments can consequently also be used for purposes other than the one for which they were originally collected or generated. Clearly, scientific research can greatly benefit from a framework that favours the reuse of data, especially if data are properly classified and structured.

Therefore, the health data regulation has been prioritized and it is not by chance that, among the sectoral data spaces envisaged by the European Union legislator, the EHDS has been the first proposal delivered[9].

Indeed, the possibility of using health data pertaining to European citizens for research purposes is perceived as a pressing issue, also in wake of the health crisis triggered by the COVID-19 pandemics. It is crucial to ensure that European researchers are not put at a disadvantage against researchers based in non-EU countries, and the existence of more pervasive and restrictive regulation may represent a concern[10]; the situation is also aggravated by the fact that to date it is not always fully clear what legal regime the secondary use of health

7 This is an initial general and a-technical distinction, useful for introductory purposes. The issue will be dealt more specifically under para. 3.

8 It should be noted that scientific research may be the primary purpose for which data are collected and processed. In such case, the research would constitute a primary use of data and therefore the relevant rules will be applied.

9 E.g., the proposal for the EU financial data space has been published more than one year after the publication of the proposal on European Health Data Space.

10 Competitiveness is a big issue for Europe. Previously, the GDPR has led to strong disinvestment in the field of medical research in favour of North America and Asia (over 10 percent). The biggest concern in the market is that there would be no point in producing health data in Europe if they couldn't be legitimately re-used. This creates a very significant damage, especially since the quality of European research is generally acknowledged.

data is subject to, since the concept of further processing that the GDPR refers to is not entirely overlapping and grey zones still exist.

Precisely to address the mentioned concerns, the proposal for a regulation on the European health data space provides for a simplified mechanism - to be managed by newly established bodies, responsible for access to health data and charged with a guarantee function – that allows health data to be accessible and marketable for research purposes: a procedure that enables the sale of health data for secondary uses, for scientific research purposes, without the need of any (direct) involvement of the data subject is envisaged[11]. However, the data that can be shared must necessarily turned into non-personal by means of anonymization or pseudonymization (if anonymized data would not serve the purposes of research) and the sharing and use must take place within safe environments through HealthData@EU.

1.3 The contribution aims to examine whether such a model outlined in the regulation will be compliant with the fundamental principles laid down in the GDPR that may appear – *prima facie* – to be in contrast, as well as to clarify what is the relationship between the health data space regulation and the GDPR.

The paper is structured as follows.

First, the current regulations in force regarding the processing of health data will be reconstructed, addressing the critical aspects and evaluating how the innovation brought by the EHDS proposal will fit into this context. Second, the (actual) potential for pseudonymization and the anonymization of health data is assessed, in order to understand the impact the new governance model could have.

11 See Chapter IV (Secondary use of electronic health data) of the EHDS proposal (specifically, Section 2 - Governance and mechanisms for the secondary use of electronic health data, Article 36 and following).

2. Health data. Defining the perimeter

2.1 As it is well known, data can be processed for a variety of purposes, each differently categorized and subject to regulations that are not entirely overlapping[12]. Such processing may differ, *inter alia*, with reference to the object to which the data refer, the purpose for which data are processed as well as the type of processing they undergo.

A literature review shows that in the context of healthcare, research and other processing of health data, multiple types of information are used[13]: electronic health records (that may include information on the symptoms declared by the patient, the results of medical exams and the relevant referral, prescriptions, etc.), claims data, omics data, clinical trials data, pharmaceutical data (including information on medicines safety). In addition to these, social media data on health, telemedicine and sensor data, information on well-being and behavioural data as well as other records of relevance to health "*such as occupational records, sociodemographic profiles or environmental monitoring data such as on pollution*"[14] can be processed.

Thus, a very wide range of information, either related to health or drawing elements about health, originating from the most disparate sources.

In other terms, the human body potentially is a system of information[15] that can be used either for healthcare or research. However, people are generally animated by a sense jealousy (and desire to confidentiality) over their health

12 There is a plurality of disciplines that refer to data, differentiated according to the type of data, the nature of the subjects to whom the data refer or by whom they are processed, and the different object of the data. Some of these regulations are complementary and apply cumulatively, while others are alternative. For an overview, see European Commission, *Study on Health Data, Digital Health and Artificial Intelligence in Healthcare*, DG Sante, July/2021, p. 185 and ff.

13 S. Marjanovic et al., *Understanding value in health data ecosystems: A review of current evidence and ways forward*, 2017, available at https://www.rand.org/pubs/research_reports/RR1972.html; European Commission, *The Use of Big Data in Public Health Policy and Research*, 2014, available at Ec.europa.eu,https://ec.europa.eu/health/ehealth/docs/ev_20141118_co07b_en.pdf.

14 European Commission, *Study on Health Data, Digital Health and Artificial Intelligence in Healthcare*, DG Sante, July/2021, p. 147.

15 I. Rapisarda, *Ricerca scientifica e circolazione dei dati personali*, in Eur. dir. priv., 2/2021, p. 301 and ff.

data, so they often they seek assurances that they will only be used for the purposes strictly necessary for their treatment and no information on their health status will be made public. Therefore, it is oftentimes arduous to obtain a patient's explicit consent to the processing of his data for research purposes, although reassured about benefits of the research and the security of the processing. Hence, the interest of public institutions and the market to the most extensive re-use of data may clash with patients' concerns (and their fundamental rights), especially with regards to data that may expose their very intimate situations and weaknesses.

2.2 A progressive development of the notion of personal data relevant to the healthcare field has been observed: from the original definition of mere medical data, the definition moved - already with European Directive 95/46/EC - to the current broader concept of health data. To the extent of governing the lawfulness of these processing, the GDPR features a broad definition of "data concerning health" as "*personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status*" (Article 4, pt. 15, GDPR); the EHDS provides an even wider notion which encompasses both personal and non-personal data[16].

Debates exist on the actual boundaries of the notion, both with regard to data that, although relating to the general health of individuals, have not been generated within a medical environment (e.g., data generated by wellness app), and to types of technologically-enabled data which do not refer directly and explicitly to a data subject (e.g., whether the so-called synthetic data – which use data collected from real patients to create a population of virtual patients with the same health characteristics – fall into the notion of health data and to

16 Indeed, the scope of health data covered by the EHDS includes also non-personal data (e.g. anonymised or aggregated data related to health and social care which may fall outside the scope of the GDPR).

what extent pseudonymization efforts allow the use and processing of health data without full compliance with the GDPR)[17].

Even outside of the perimeter of healthcare purposes, health data may acquire economic value for several uses and applications: some forms of reuse of data relate to patient profiling (for diagnostic or prevention purposes) as well as to scientific research purposes (prospective or retrospective)[18]; additionally, they may be used also for purely commercial purposes if a legal basis for processing exists[19]. Alongside their legal soundness, the ethical aspects of such transactions should also be considered.

2.3 To address these multiple concerns, in the context of the regulations already in force, the processing of (personal) health data is subject to special rules that partially derogate from those applicable to personal data in general. These rules are, in principle, more stringent precisely to meet the increased demands for protection, but at the same time they aim to make easier those processing capable of producing positive societal effects.

The proposed EU Data Health Space, albeit frequently referring to the GDPR and to its definitions, often (e.g., in relation to types of data, processing and entities having a role in the chain of data collection, use and exploitation)

17 The concept of electronic health data refers to both personal and non-personal data concerning health and genetic information as well as the relating "determinants".

18 I.V. Pasquetto - B.M. Randles - C.L. Borgman, *On the Reuse of Scientific Data,* Data Science Journal 16(8), 2017, p. 1–9. Traditionally, in Italy the observational retrospective studies faced many obstacles with reference to the interpretation of norms provided by the Italian Data Protection Authority (see Opinion pursuant to Article 110 of Italian Privacy Code and Article 36 GDPR - 30 june 2022) that required a specific new consent by the data subjects in order to legally use such health data or alternatively, the obtainment of a preemptive authorization on the study by the territorial ethical committee and by the Data Protection Authority. The recently approved reform of Article 110 is aimed precisely at overcoming this shortcoming.

19 On the debate on marketability of data, see V. Janeček-G. Malgieri, *Data Extra Commercium*, in S. Lohsse, R. Schulze and D. Staudenmayer (edited by), Data as Counter-Performance—Contract Law 2.0?, Hart Publishing/Nomos 2020, pp. 93-122; G. Alpa, *La proprietà dei dati personali*, in N. Zorzi Galgano (a cura di), *Persona e mercato dei dati. Riflessioni sul GDPR*, Milano, 2019; H. Zech, *Data as a tradeable commodity*, in A. De Franceschi (a cura di), *European Contract Law and the Digital Single Market. The Implications of the Digital Revolution*, Cambridge, 2016, pp. 51 and ff.

makes use of different nomenclatures[20], most of them introduced by the Data Governance Act[21] and the Data Act[22]. Clearly, there is an effort on creating a common 'platform' (in terms of applicable rules, technology and governance) for the use and valorisation of (health) data, trying to overcome the existing regulatory inconsistencies.

Significant changes were made to the first versions of the proposal to address the identified concerns[23], nonetheless misalignments still exist between some of the definitions included in the EHDS proposal and the ones featured by the GDPR[24]. In the present analysis, these elements and their consequences will be highlighted.

3. The state of the art in health data processing

3.1 To date, the main regulatory source on the processing of health data (at least those that qualify as personal data) is the GDPR. In fact, Article 9 regulates the processing of 'particular' types of data, among which health data falls. The inclusion of health data in this category is justified as their specificities require greater attention as well as ease when it comes to processing for public interest purposes. While it does not imply a complete ban on their processing, however, subjects it to the specific conditions listed under Article 9 (2). The consent-centric model has been *de facto* superseded, to the extent that the regulation provides for a variety of legal bases for processing (other than explicit consent of the data-subject), including some that rely on

20 D. Amram, *Comparing EU initiatives on data,* cited, p. 7.

21 Regulation (Eu) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).

22 Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act).

23 Most of the weaknesses of the proposal, pertaining to data protection, were raised in the EDPB-EDPS, Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space, adopted on 12 July 2022.

24 E.g. the definitions of primary and secondary use set forth under Article 2 (2) (d) and (e) EHDS.

the existence of a public interest to the processing[25]. In this context, scientific research purpose already constitutes a valid legal ground for processing as long as it entails a public relevance. At present, simplified schemes aimed at making scientific research practicable even in the absence of an express consent from the data subject exist, yet they are subject to stringent limitations and no dedicated infrastructure aimed at making them possible on a large scale is available.

3.2 Thus, in the GDPR there is already a clear trace of the balancing between the existing interests underlying health data processing: interests of the data subject to data protection and control over its data, on the one hand, and public interest to public health and scientific research, on the other[26].

In this respect, the GDPR proves to be research friendly[27] establishing a presumption of compatibility[28] of the further processing for scientific research purposes (Art. 5(1)(b))[29]. In fact, when it comes to secondary use for scientific research, the Regulation states that the (further and secondary) scientific

25 Legal models that rely (exclusively or mainly) on the consent of the data subject as the legal basis for processing pose problems for scientific research: not only because of the difficulties, in some situations, in obtaining express consent, but mainly because the research activity could give rise to the re-use of the data and their further proliferation. This is more evident than ever due to the current mass digitization W. Ricciardi, *Assessing the impact of digital transformation of health services: Opinion by the Expert Panel on Effective Ways of Investing in Health (EXPH),* 2019, 29 European Journal of Public Health Supplement 4, p. 185–769.

26 Notwithstanding the applicability of the general principles of purpose limitation and data processing minimization, further processing may only be considered legitimate if the controller has carried out a compatibility assessment (Article 6(4) specifies what are the parameters to be taken into account), on the basis of the data subject's consent or on specific rules set forth by European or national laws.

27 On the friendliness of GDPR for scientific research, see G. Comandè - G. Schneider, *Can the GDPR Make Data Flow for Research Easier? Yes It Can, by Differentiating! A Careful Reading of the GDPR Shows How EU Data Protection Law Leaves Open Some Significant Flexibilities for Data Protection-Sound Research Activities*, *Computer Law & Security Review* 41, 2021, 105539, available at DOI: 10.1016/j.clsr.2021.105539.

28 Similarly, the Convention no. 108 of the Council of Europe already provided for (similar) presumptions of non-incompatibility when further processing carried out was aimed at scientific purposes, as long as it was backed up by adequate safeguards.

29 Article 5 (1) (b) GDPR and Recital 40 provide that further processing for research, statistics and historical purposes may be considered as "not incompatible", allowing therefore an additional processing with a non-incompatibility presumption.

research finality is to be deemed generally compatible with the initial purposes – and related legal basis – as long as the processing is conducted in compliance with the criteria laid down in Article 89[30].

When none of the previous is applicable, a specific compatibility test (Article 6 (4)) aimed at carrying out the processing may still be done[31]. However, in this case the data subject must be notified according to Article 13 (3) or 14 GDPR)[32].

Though, a secondary processing for scientific research purposes is not necessarily (*de jure*) compatible with the initial legal basis, regardless of the interests pursued. In fact, any such reading would be in contrast with the interests of the data subject and would contradict the provisions of Recital 33, which requires that - in the context of scientific research - specific areas of study are still to be identified[33].

Already from this debate, the relevance (and, at the same time, the applicative and technical/legal difficulties) of re-use for research purposes can be grasped. In order to better determine the concrete scope of applicability of the rules examined, however, it is necessary to consider one point in more detail.

---

30 G. Schneider, *Disentangling health data networks: a critical analysis of Articles 9(2) and 89 GDPR*, International Data Privacy Law, Volume 9, Issue 4, November 2019, pp. 253–271, available at https://doi.org/10.1093/idpl/ipz015

31 Article 99 of the Italian Privacy Code states that, for the purposes of scientific research, personal data whose processing has ceased (for any reason) may be retained or transferred to another data controller, in compliance with the guarantees provided for in Article 89 GDPR.

32 In view of the obvious difficulties that sometimes exist in providing such information, Article 14 (5) GDPR provides for a special exemption from the obligation to provide information on processing when providing it would be impossible, would involve a disproportionate effort (in terms of high costs, money, time) or would risk impairing scientific research by jeopardising its objectives. For these reasons, the EHDS provides for further simplification in this regard.

33 In the same vein, the Article 29 Data Protection Working Party, *Opinion 3/2013 on purpose limitation*, 00569/13/EN WP 203, 2 april 2013, p. 28 ff. reiterated the need to identify a specific legal basis for processing, keeping the compatibility test and the legal basis for processing separate.

3.3 The case law[34] and the EDPB[35] have repeatedly stated that the legal bases provided under Article 9(2)(j) GDPR cannot be used when the research is motivated by lucrative purpose. Such interpretation seems to link the benefit of a more advantageous legal basis for the processing of health data to the circumstance that the beneficiary is a public body, entrusted with the care of the public interest, rather than solely to the objective pursuit of the public interest objective underlying scientific research (which is still devoted to the improvement of scientific knowledge, albeit with the ultimate profit motive of the company promoting it)[36]. Widening and enhancing the research purpose envisaged by the GDPR would allow to legitimise further secondary processing pursuant to Article 9(2)(j) fostering cooperation between hospital, public institutions and private research organisations. It would facilitate the valorisation and re-use for scientific purposes of health data held by public bodies even when it is carried out by private parties performing 'private' research[37]. Precisely the objectives pursued by the EHDS proposal.

Accepting the more restrictive view, instead, such further processing would necessarily require a new consent by the data subject[38], clearly harming the practical viability of secondary processing. Hence, especially in these circumstances, it would greatly benefit research allowing forms of broad consent that would enable the data holder to carry out a number of different processing – even not strictly related one to the other – without the need of

34 Among others, Italian Supreme Court of Cassation, Sec. I, 7 October 2021, no. 27325.

35 European Data Protection Supervisor (EDPS), *A Preliminary Opinion on Data Protection and Scientific Research*, January 2020.

36 Accepting a broader interpretation of the concept of scientific research - also from an objective point of view - may grant "full freedom to operate" over the data to research bodies.

37 M. Mostert et al., *From Privacy to Data Protection in the EU: Implications for Big Data Health Research*, in European Journal of Health Law, 25, 2018, p. 52 highlights that in health studies, to enable the best progress and unlock further research perspectives, together with the circulation and use of data already available, it would be necessary to allow to cross-reference them with those in further databases (e.g. data from pathological anatomy archives, biobanks or even social networks).

38 In specific cases (see 36 GDPR), a specific preemptive authorization by the national data protection authority could similarly legitimize such further processing.

new contacts with the data subject. Are those models – or forms of implicit consent – permitted as of now? Already before the entering into force of GDPR, WP29 expressed concerns in relation to forms of broad consent[39] as a legal base for the processing of health data. The feasibility of models of tacit consent to further secondary processing, as well as privacy disclosures with 'open' consent when the further use contemplated is motivated by purposes of medical progress, and of other models in line with the rationales underlying the European data space should be investigated more in depth[40]. However, these arrangements should be examined with caution as they would lead to a derogation to the core principles of data processing minimisation and determinateness of the uses to which consent is given[41].

Some insights in this regard can be understood from the extensive debate that exists relating biobanks. Those by their nature are repositories of data set up for future research, unspecified (or at least not fully specified) at the moment of their establishment and delivery of biospecimens and relating data. In this field, even international ethics norm governing health research are increa-

39 Article 29 Data Protection Working Party, Opinion 03/2013 on purpose limitation, 00569/13/EN WP 203, Adopted on 2 April 2013.

40 It is significant, however, that some jurisdictions – building on Article 36 GDPR – have already developed solutions to allow the use of data for research purposes without the consent of the data subject, at the request of the data controller and subject to the favourable opinion of a special commission (which first assesses the merits of the purposes pursued). For example, Irish law provides for a similar mechanism. See D. Amram, *The "Accountable Ulysses". Research and Health Data Protection: How to Harmonize at European Level despite of the GDPR and the Interpretations offered by the Irish, Belgian, Spanish, and Italian systems*, Rivista Italiana di Medicina Legale (e del Diritto in campo sanitario), fasc.1, 1 February 2019, p. 215. Similarly, Article 110 bis of the Italian Privacy Code provides for the possibility of reusing personal (health) data, even in the absence of the consent of the data subject, provided there is a prior authorisation by the Data Protection Authority (and provided specific safeguards are implemented for the minimisation of the processing and for the anonymization of the data). with the recent reform of Article 110 of the Italian Privacy Code, the utilisation of such data for scientific research that complies with deontological rules is made even smoother, with no need for prior authorisation by the Authority.

41 Another option theoretically viable is to presume the legitimacy of the further processing, via forms of presumptions of consent to secondary use (granting protection to the data subject position through the option of opting out).

singly approving forms of broad consent[42]. The main applicative question is how to ensure that the use of such data would be compliant with the type of uses already envisaged at the moment of collection. Although in theory the model outlined in that sector could offer solutions to be applied more broadly to health data, it does not seem however to strike a balance that can be applied to all forms of secondary use of health data, which are manifold and present multiple protection requirements.

4. Secondary use and its borders. Speciality or complementarity between health data regulation and GDPR?

4.1 The secondary use of health data has many practical applications, yet it raises legal issues of considerable complexity.

Even if it is a widely used concept, the GDPR does not define nor make use of the term 'secondary use' and the notion further processing it features does not appear to be completely overlapping. The latter exists whenever the processing is made for a purpose different to the one the data were collected for, and in this context no relevance is given to data lifecycles as discriminating factor. Although the further processing is not banned entirely, it is considered exceptional and subject to penetrating rules.

Secondary uses, instead, are generally often defined as any use of data, different from the one for which data have been collected or generated[43]. In

42 E.g., both World Medical Association and the Council for International Organizations of Medical Sciences/World Health Organization (CIOMS/WHO) in 2016 have agreed for the use of such model. The nature of biobanks is not well suited for specific consent as a legal basis for processing because the activities they carry out require continuous access to information, biosamples and the individuals themselves who have participated in the research, who are often followed up and recontacted by the biobank for collection of further data, submission of questionnaires, measurements, or even new samples. Therefore, if consent is to be used as a legal basis for processing, it is inconceivable to subordinate it to the principle of specificity of use. Alternatively, other legal bases that are independent of consent could be used.

43 World Health Organization, *Meeting on secondary use of health data*, 13 December 2022, available at https://www.who.int/europe/news-room/events/item/2022/12/13/default-calendar/meeting-on-secondary-use-of-health-data#:~:text=Secondary%20use%20of%20health%20data%20is%20the%20processing%20of%20health,of%20a%20service%20or%20product. for example states that "*Secondary use of health data is the processing of health data for*

fact, data can serve additional purposes and therefore may be exploited by the data holders or by third parties to this end.

Secondary use of health data unlocks greater efficiency because of economies of scale and scope made possible by aggregation of different health datasets that "*usually yields more (accurate) insights compared to the separate analysis of fragmented datasets*"[44]. Additionally, it proves functional to increase transparency in the provision of health services to the market, as the analysis of a wide range of data has great potential in comparing the efficiency of the different healthcare services. This has an impact both on service providers, who can better modulate their offer, and on patients themselves, who can more consciously choose the most suitable services, and possibly changing their service provider, without losing their health data already collected (clinical analyses, diagnosis, etc.) by using the data portability tool. As for research, building on the actual evidence of a therapeutic intervention avoids duplication of studies and complies not only with efficiency goals but also with the ethical requirement of preventing patients being subjected to studies whose outcome is already known, as it is contained in the available data or could be inferred from them[45].

The EHDS proposal innovates the examined landscape, introducing a definition of secondary use[46] that reflects the novel paradigm proposed.

In the context of the impact assessment conducted by the European Commission on the EHDS the secondary use (or reuse) of health data is

purposes other than the initial purposes for which the data were collected."

44 Idem, p. 11 which in turns quotes B. Carballa-Smichowski et al., *Economies of scope in data aggregation: evidence from health data*, 2022, available at https://www.researchgate.net/publication/365276562_Economies_of_scope_in_data_aggregation_evidence_from_health_data. However, it must be mentioned that the characteristics of these datasets are such that, upon reaching a given (significant) size and level of accuracy, further expansion may bring only a minimal benefit in terms of insights on patients; it must also be considered that the size and characteristics of the dataset to be used to obtain useful medical insights may vary (considerably) depending on the complexity of the data and diagnostic questions.

45 These (further) uses of data which prove to have so many positive points, however, raise criticalities especially from a data protection and cyber risk standpoint.

46 EHDS Proposal, Article 2(2)(e).

regarded as the use of "*individual-level, personal or non-personal health data or aggregated datasets, particularly data generated during healthcare provision with the purpose of supporting research, innovation, policy making, regulatory activities and other uses, such as healthcare delivery to a patient, based on the data concerning other patients*"[47].

Within the Regulation, therefore, the provisions on the secondary use of data, which aim to broaden its scope and make its procedures smoother, refer to the defined concept. However, in the broader area of data processing there is no full consensus on the extension of the concept, that often is used as a synonym for "data reuse" and "repurposing"[48]. These, though, are not entirely alike. Nonetheless, such terms – and other semantically related that are generically used to refer to such processing – highlight some of the core features that in additional processing of health data exist (*i.e.*, time interval separating the data generation and first to the subsequent uses, possible non correspondence of data user and original data controller, identifiable different purpose for data use).

Among the possible classifications useful to draw a line between primary and secondary use of data, the definitions established in the biomedical research domain (and the differences with the concept of further processing featured by the GDPR) can be recalled. In that field, the difference between secondary use and uses that fall outside such scope are attributable to whether

47 European Commission, *Impact assessment report Accompanying the Document, Proposal for a regulation of the European Parliament and of the Council on The European Health Data Space*, p. 12. In the background of the EHDS proposal, it has been emphasized that incentivizing the re-use of health data will be valuable, as it can greatly benefit the development of new products, with an expected increased value "estimated at around EUR 25-30 billion at present, expected to increase to around 50 billion in 10 years".

48 The lack of consensus on the definition of secondary use (and clear demarcations between this concept and others close to it) created barriers to cross-border data sharing. Additionally, lack of clarity on the meaning and its scope (and, even more, its inconsistent use) can determine significant challenges when obtaining consent as well as difficulties to interpret what individuals have consented to. Among the negative consequences, difficulties to access certain types of data (e.g., genomic data or data from certain subjects) due to overly cautious and risk-averse behaviour by data controllers and underutilization of health data as a resource for secondary use. This leads to inefficient use of resources, human and financial, due to the need of re-collection of the same health data for other purposes.

the processing for an additional purpose concerns an entire data lifecycle *(i.e.*, all the possible activities and processing from data generation to deletion) or not. Instead, according to a different criterion (a controller focused view), the discrimen is to be found in the existence of a distinct phase within the data lifecycle: according to this approach - focused on the identification of the controller - each phase within the data lifecycle begins when a controller collects personal data, either directly from the data subject or, in the case of existing data, from another source, and ends with the realisation of the purpose for which that controller collected the data[49].

In such a fragmented landscape, the EHDS has adopted a more clear definition to provide greater certainty, and with the aim of favouring the re-use of health data, being aware of its potentialities: indeed, secondary use of health data is suitable for supporting, among the others, health research, innovation, policymaking, regulatory and personalised medicine[50] and it may streamline the process (and make it less costly) for the development of new drug medicines and protocol for medical procedures[51].

4.2 Notwithstanding the different nuances in the understanding of what is secondary use and what still falls under primary use (or even qualifiable as something else, for example being outside the scope of personal data), the secondary use concept has a specific relevance in the field of health as it unlocks the possibility of providing value-added services on the market as well as potentialities for research[52] (even though often the potential of secondary

49 Hospitals that administer medical examinations and tests may use the data collected both for diagnostic purposes (primary use) and for researches not linked to the cure of that specific illness (a secondary use, as this processing is unrelated to the main purpose the data were collected for). See idem, p. 137. The same applies when the research additional processing is not carried out directly by the original data holder but by a third party.

50 J.S. Marcus et al., *The European Health Data Space*, cited, p. 20.

51 As well as public health decision making. During the COVID-19 pandemics the use of health data proved to be essential.

52 C. Safran et al., *Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper*, Journal of the American Medical Informatics Association 14(1) 2007, pp. 1–9; I. Danciu et al., *Secondary Use of Clinical Data: The Vanderbilt*

use is not sufficiently exploited because of the privacy and data protection concerns that it raises)[53]. Therefore, the need of establishing a functioning system for authorizing such processing appears to be essential. Clarifying its scope and assessing whether secondary use of health data constitutes – always, sometimes, when? – a form of further processing under GDPR[54] is also crucial, as uncertainty represents a relevant issue exposing researchers to non-compliance risks. Whereas Recital 37 of the EHDS proposal states that the mechanism provided for in Art. 34 EHDS (which envisages the purposes for which health data may be processed for secondary use) constitutes a legal basis legitimising the processing of the data under Art. 6 and 9 GDPR, it doesn't fully solve the problem.

In general terms, the EHDS has not cleared up all the sticking points with the GDPR and the applicative difficulties arising from uneven regulatory regimes. Earlier during the *iter legis* procedure, the joint opinion issued by the EDPB-EDPS pointed out some of the most relevant criticalities of the proposal, when it comes to data protection and rights of data subjects[55]. It should be appreciated that most of those were addressed and sorted out in the course of the negotiations on the text. Nevertheless, some outstanding questions remain.

Approach, Journal of Biomedical Informatics, Special Section: Methods in Clinical Research Informatics 52, 2014, pp. 28–35, available at DOI: 10.1016/j.jbi.2014.02.003; S.M. Meystre et al., '*Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress*, Yearbook of Medical Informatics 26(1), 2017, pp. 38–52, available at DOI: 10.15265/IY-2017-007.

53 L.O. Gostin - S.F. Halabi - K. Wilson, *Health Data and Privacy in the Digital Era*, Journal of the American Medical Association 320(3) (2018) pp. 233–234, available at DOI: 10.1001/jama.2018.8374.

54 R. Becker et al., *Secondary use of Personal Health Data: when is it 'Further Processing' under the GDPR, and What Are the Implications for Data Controllers?*, European Journal of Health Law 30, 2023, pp. 129–157, available at https://ssrn.com/abstract=4070716 or http://dx.doi.org/10.2139/ssrn.4070716.

55 EDPB-EDPS, *Joint Opinion 03/2022 on the Proposal for a Regulation on the European Health Data Space*, adopted on 12 July 2022. Comments on the joint opinion EDPB-EDPS were offered by G. 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, 8:3 Eur Data Prot L Rev 398; V. Cimina, *The Proposal for a European Health Data Space: between pursued objectives and data protection challenges*, ERA Forum (2023) 24, pp. 343–359

Although the draft Regulation now expressly emphasizes the complementarity between the rules on the processing of health data contained in the EHDS and those of the GDPR[56], their interplay is not always clear from the point of view of their coherence and/or speciality. In fact, the relationship between the two pieces of legislation lies in coexistence but inconsistency, both in terms of definitions and data governance models. This results in a diminished protection of data subjects.

This reading key will therefore be used to analyze the provisions of the EHDS proposal.

5. (continues) The European health data space and its simplified mechanism

5.1 The European Data Strategy acknowledged the value that data has and possible further uses they may unlock. Therefore, the EU Commission promoted a legal framework aimed at favoring the development of technologies, preserving the European values and protecting data subjects' rights; at the same time promoting the data circulation and forms of processing aimed at exploiting the value of data[57]. In a regulatory environment marked by the presence of horizontal disciplines, such as the Data Act and the Data Governance Act, the EHDS builds on their definitions and compliment such regulations providing specific rules for health data: it covers "*standards and specifications for providers of data intermediation services in the health sector, minimum technical requirements for the portability of health data, criteria for security of data for bodies dealing with data altruism*" [58] .

56 Under Art. 1, para. 4 EHDS it is expressly stated that the Regulation *"shall be without prejudice to other Union legal acts regarding access to, sharing of or secondary use of electronic health data, or requirements related to the processing of data in relation to electronic health data, in particular Regulations (EU) 2016/679 […]".*

57 European Commission, *A European Strategy for data*, Shaping Europe's digital future, June 2023, available at https://digital-strategy.ec.europa.eu/en/policies/strategy-data.

58 Each data space will feature distinctive and sector-specific rules and models of governance. It may also feature different technological characteristics. In such regard, see V. Cocca, *European perspectives: the creation of data spaces in the health care sector*, in Data Valley White Paper e-Health Data Sharing Best practices and solutions for data sharing, anonymization, and data lake creation with health data, October 2021.

The proposal pursuit several objectives, listed under Article 1. While strengthening the rights of natural persons in relation to the availability and control of their electronic health data (lett. a) and laying down rules for developing electronic health records systems (lett. b), it aims to provide rules and mechanisms supporting the secondary use of electronic health data (lett. c), establishing a mandatory cross-border infrastructure for the secondary use of electronic health data. Thus, it impacts on both primary[59] and secondary use of health data and aspires to improve both.

With the aim of reaching a balance among the conflicting interests (and concerns), the proposal features a complex mechanism, whose functioning is guaranteed by public institutions - charged with management and guarantee functions - and where all the involved players (the health data access bodies, data holders, data users and data subjects) are tasked with specific duties and are right-holders.

It tackles the identified issues of public health, efficiency and data protection proposing a new governance structure - technological and institutional - for the control on health data (enhancing the tools to raise awareness on the use of data and the subsequent possibility to exercise the data subject rights), data circulation (data portability rights as well as data sharing mechanism), possible uses in diagnosis, healthcare and research enabled by interoperability.

More specifically in relation to secondary use, it provides that electronic health data[60] – collected and processed in various contexts with the support of public funding (EU or national) – should be made available by data holders[61] to

59 The primary use is defined under Article 2 (2) (d).

60 While Article 33 identifies the (minimum) categories of health data that should be shared, under Article 34 the purposes that may justify the secondary use are listed (among them also the research purpose, Art. 34 (1)(e). It should be emphasized however that although the proposal intends to shed light on many problematic profiles to date, it does not completely clear the field of questions. On the contrary, it sometimes raises new questions that it is hoped will be resolved as the legislative process continues. For example, under Article 34(1) of the Proposal, not all the types of secondary use featured are homogeneous and they would fall under different hypothesis of exceptions under Article 9(2) GDPR.

61 The GDPR itself provides the technical impossibility as a limitation to the transferability of personal data. In fact, it states that the data controller must be able to directly transfer

potential data users (through health data access bodies[62]) with the declared goal of maximizing the impact on research, innovation, patient safety or policy making, generating thus a benefit for the society[63].

However, the pursuit of such objectives raises various concerns, especially related to the processing of particularly sensitive data, to which unrestricted access cannot be provided to data users and security conditions must be ensured: principles such as "*privacy by design*"[64] and "*bring questions to data instead of moving data*" should be implemented[65]. Therefore, the mechanism envisaged by the EHDS features some precautionary measures and among them notable are the ones of structural nature. In this respect, all operations should take place through HealthData@EU and within a secure data processing environment where certain key principles are implemented (identification of user, minimization of use of data accessed, compliance with security measures). This provides greater reliability to the whole mechanism, but at the same time, seems to legitimise - in the approach of the proposal - derogations (in some points) from the GDPR rules, in the sense of greater support for data circulation and reuse.

The functioning of the EHDS secondary use mechanism is guaranteed by the creation of Health data access bodies (Article 36) designated by each member state and responsible for granting access to electronic health data for

---

portable data to another data controller indicated by the data subject, if technically possible. In the context of the new proposal, HealthData@EU is the infrastructure that makes this possible and thus justifies exemptions in the EHDS proposal to the GDPR.

62 It should be however mentioned that under Article 49 of the EHDS proposal when an applicant requests access to electronic health data only from a single data holder in a single Member State it may file the request directly to the data holder and the data holder may grant the access complying with the relevant conditions set forth under Article 46.

63 In countries where private institutions play a fundamental role in healthcare, the obligation to data sharing should be extended to those subjects as well. However, the sharing obligation does not provide for reciprocity. In fact, private entities and pharmaceutical companies benefit from access to such data to be employed in their research, but are not in turn obliged to share the health data they hold.

64 On the implementation of the privacy by design in the health field, see G. Bincoletto – P. Guarda, *A proactive GDPR-compliant solution for fostering medical scientific research as a secondary use of personal health data*, Opinio Juris Studies in Comparative and National Law, no. 1/2021.

65 See Recitals 40 and following of the EHDS proposal.

secondary use. Those bodies will have a *de facto* role of gatekeepers of data subjects' health data being entrusted with 'an intermediary role that is twofold'[66]: on the one hand, to protect the data subjects from the re-identification risk, while on the other hand, to provide technical assistance to data users in selecting a suitable dataset according to the approved request of access, verifying the compliance of all the actors involved with the relevant rules - giving also "*health data users and holders the opportunity to reply to any findings and to remedy any infringement*"[67] - and, if necessary, impose sanctions[68]. According to Article 37 and following, these bodies are entrusted with many functions as will have, e.g., to decide on data access applications, to provide access to electronic health data for the purposes set out in Article 34, maintain a management system to record and process data access applications, facilitate cross-border access to electronic health data for secondary use hosted in other Member States through HealthData@EU and cooperate closely with each other and with the Commission. Morover, they will be in charge to make public a national dataset catalogue that shall include details about the source and nature of electronic health data and the conditions for making electronic health data available[69].

66 D. Amram, *Comparing EU initiatives on data,* cited, p. 9.

67 As provided for by Recital 48 of the EHDS proposal.

68 There is however the risk that the role attributed to data access bodies in the access mechanism for secondary use envisaged by the EHDS will lead to a subtraction of the competence of the data protection authorities. The scrutiny currently possible under Art. 36 GDPR would be depleted, granting the access bodies only the responsibility to express opinions on the scientific research carried out and thus on the usability of the data, allowing or not allowing access (purchase). To cope with these criticalities, the current article 11a EHDS expressly acknowledges the competence of the data protection supervisory authorities both "*for monitoring and enforcement of Regulation (EU) 2016/679*" and "*for monitoring and enforcement of the application of Articles 8a to 8f and Article 8h*" EHDS.

69 This model is in line with the FAIR principles (findable, accessible, interoperable, and reusable), that ensure data quality also through standards and allow the re-use of secondary data and unlock possibilities of developing data driven applications. See A.F. Näher et al., *Secondary data for global health digitalization*, Lancet Digit Health 2023; 5: e98.

5.2 The EU datasets catalogue should minimise the administrative burden for the data holders and other database users; be user-friendly, accessible and cost-effective, connect national data catalogues and avoid redundant registration of datasets. And Member states should ensure that data catalogues kept at a national level are interoperable with existing dataset catalogues from European research infrastructures and other relevant data sharing infrastructures[70]. Furthermore - and this is essential to ensure truly valuable data for use in scientific research – the data included in the catalogue shall ensure a minimum level of data quality and the dataset shall be sufficiently described, so as to enable potential data users to figure out which of those information might be useful for their research.

In this context, therefore, data users who wish to access the data must make a request under Article 45, stating, *inter alia*, what data (in the catalogue) they intend to access, what use they intend to make of it, what security measures they will employ. Once the request has been assessed and accepted by the national Health data access body[71] (that has ordinarily a term of 2 months to issue the decision) the data are immediately requested by the body from the data owner, in order to make them available to the data user.

The authorisation issued by the national body may benefit from mutual recognition by the other national bodies responsible for accessing the data. However, if a single data space is to be created, there should be a real obligation for automatic recognition.

As to the conditions of access to data by the data users, it may require the payment of fees under Article 42[72], the amount of which is determined on the basis of objective criteria. However, it is worth clarifying the economic *rationale* of this provision: the activity of data collection and storage is

70 EHDS Proposal, Recital 60.

71 Access to health data by EU and national institutions, bodies or organisations are not subject to authorisation (see e.g. EMA).

72 The provision of fees for re-use of data is a possibility already mentioned under article 6 DGA.

expensive; even more so is compliance with the obligations to structure electronic health records according to certain formats, to be made available on the platforms envisaged by the EHDS proposal. While access to such data as primary use does not envisage any form of payment (and thus remuneration for the person who generated it), secondary use is subject to specific charging. Such a pricing mechanism does not seem likely to generate considerable revenues for data holders, but could contribute to a sustainable organisational transformation and to the compliance costs that organisations will have to bear in order to fully implement the EHDS model.

It should be emphasised, though, that the possibility for data users to select the desired data from a catalogue does not pave the way for indiscriminate dissemination of all personal health data. Pursuant to data minimization principle, the health data access body shall ensure that access is only provided to requested electronic health data relevant for the purpose of processing indicated in the data access application by the data user and in line with the data permit granted. The health data access bodies shall provide the electronic health data in an anonymised format, where the purpose of processing by the data user can be achieved with such data, taking into account the information provided by the data user.

Where the purpose of the data user's processing cannot be achieved with anonymised data, taking into account the information provided by the data user, the health data access bodies shall provide access to electronic health data in pseudonymised format (and the information necessary to reverse the pseudonymisation shall be available only to the health data access body and not to the data users).

Hence, the circulation of health data is encouraged, and as long as this takes place for research purposes and by structured entities, the circulation of

pseudonymised data is also allowed in cases in which anonymised data[73] would not meet research needs.

In order to address criticism about the watering down of data protection in favour of data circulation, the Regulation now provides for the right of the data subject to opt out from secondary usages of his health data. Although he has no effective control over actual uses and decision-making on sharing, he may decide to exclude his data from such a mechanism. However, this is a reversible decision, reflecting the logic of favouring maximum re-usability of data.

A flaw in the system - considered from the point of view of the protection of the data subject – is that Member States are entitled to adopt laws intended to override the eventual opt-out by data subjects to enable secondary uses by public bodies, which are assigned with the pursuit of public interests deemed particularly meritorious.

An opt-in system is implemented only for data deemed extremely sensitive, in respect of which the balance between data subject protection and circulation has therefore been settled in favour of the former.

6. What use for anonymized and pseudonomyzed health data

6.1 At this point, we should better investigate to what extent a balance is achieved between data circulation and data protection. This requires dealing more specifically with the concepts of pseudonymisation and anonymization as the EHDS proposal, with regards to the secondary use for research purposes, builds on such notions.

Such processing - whose performance is mandated to the data holder or health data access bodies -allows for a wider circulation of data, regardless of the express consent of the data subject.

73 It has been proved that some form of anonymization still preserves many of the features of data essential to carry out high quality research. See M. Donnelly - M. Mcdonagh, *Health Research, Consent and GDPR Exemption*, in *European Journal of Health Law*, 2019, 26 (2), p. 97 ff.

Pseudonymisation is a sufficiently clear concept (at least on paper) as the GDPR, under Art. 4, defines it as a form of "*processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information […] kept separately and is subject to technical and organisational measures*". Conversely, it does not provide a specific definition of anonymization. The Recital 26 merely refers to anonymous information as "*information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable*."

The main difference among anonymization[74] and pseudonymization lies in the reversibility of the de-identification operation, and the consequent possible re-identification of the data subject, which determine the applicability (or not) of the data protection rules. Simplifications resulting from the non-applicability of GDPR rules are not, however, without drawbacks: the data thus processed are depleted and the range of uses that can be made of them is reduced (also with regards to the secondary use for scientific research purposes). In fact, anonymised data can potentially be inaccurate: excessive anonymisation may reduce the quality, usability and reliability of the data to the point of making them of little use for the purposes pursued. Indeed, in the context of research, it is often important to be able to make connections between data and subjects, particularly in the context of correlation studies.

As discussed, the model for the secondary use of health data outlined by the EHDS heavily relies on the use of pseudonymisation and anonymisation techniques, insofar as it subordinates data users' access to health data for secondary use to such processing. The question that arises is whether the application of such techniques sufficiently protects the data subject from whom the

74 The anonymization outcome may be obtained with different technological tools and techniques. Among them, for example, the elimination of certain relevant parameters from the dataset, randomization by means of 'noise' addition, generalization.

data (later anonymised) originates, who thereby loses all ownership and control.

The risk of re-identification sometimes remains even after anonymisation procedures as certain categories of electronic health data can remain particularly sensitive even when they are in anonymised format. For example, a residual risk of re-identification (also through techniques not available at the state of the art, but that may be developed later on) remains for example in relation to rare diseases where only a limited number of cases exist. On the other hand, the lack of homogeneous guidance across the Union ultimately makes the procedure more or less burdensome and effective in different contexts.

These findings seem to undermine the interplay of interests pursued (and allegedly achieved) by the EHDS proposal.

Moreover, a distinction should be done between data that are absolute anonymised (with the removal of the relevant "elements", so that re-identification of the data subjects is impossible) and *de-facto* anonymisation that occurs when the re-identification is still technically possible, but only with an unreasonable effort in terms of time, cost and manpower. It is relevant also for regulatory purposes as only the absolute anonymized data can be accessed both by individuals and organizations, while the *de-facto* anonymized ones can be made available only to scientific institutions which pursue a research purpose. This point shows the regulator's awareness of the weakness of the model in guaranteeing full protection of the data subject. Nevertheless, no definitive solutions have been proposed to date, but risk mitigation elements have been introduced (such as, for instance, the obligation - mentioned above - to perform activities within secure environments).

As a consequence of the uncertainty on the borders of the concepts[75], overly risk-averse behaviours (e.g., considering all data as personal data for the purpose of identifying the applicable rules) are often kept by research institutions leading to reduced re-use of health data and a subsequent slowdown on innovation. On the other hand, over-anonymisation aimed primarily at escaping the application of the burdensome GDPR rules "*can significantly reduce the quality, usability and reliability of health data*"[76].

In light of this - and also aiming to align the new health data governance model more closely with the objectives of the GDPR - a possible way of enhancing the current proposal would consist in including a clear guidance on the interpretation of the concept of 'anonymisation', as hinted also by Recital 49 EHDS. Similarly, it would be useful to outline a common European interpretation of what constitutes 'pseudonymisation', identifying "*the degree of separation needed between the re-identification key and the data user for data to be considered pseudonymized*"[77].

The current situation creates a number of obstacles for research friendly environments that risk jeopardising the data subject's control over personal data also when no such need arises for research purposes, and exposes researchers to legal risks of non-compliance, as the actual extent of the relationship between Article 34 EHDS and Articles 6 and 9 GDPR mentioned above is still unclear. Furthermore, the regulatory uncertainties also pose

75 In fact, some countries apply a stricter definition of 'sufficient anonymisation'. Of re-identification (e.g., due to the rarity of their illness). There are also evident difficulties in following the patient in its healthcare, when each provider sticks to different interpretation of the 'sufficient anonymisation' concept. Even more problematic the fact that each institution may follow different methods to achieve anonymization as there is a lack of guidance on how to achieve it (at least with reference to some categories of personal health data, e.g. medical images). See M. Shabani - L. Marelli, *Re-Identifiability of Genomic Data and the GDPR: Assessing the Re-identifiability of Genomic Data in Light of the General Data Protection Regulation*, EMBO Reports 20(6) (2019) e48316, available at DOI: 10.15252/embr.201948316.

76 TEHDAS - Joint Action Towards the European Health Data Space, *Report on secondary use of health data through European case studies*, 28 February 2022.

77 TEHDAS - Joint Action Towards the European Health Data Space, *Report on secondary use of health,* cited, p. 14.

applicative challenges. For example, there are difficulties in making data pools interoperable due to different standards and methods of pseudonymization and the provision of different procedures and of specific safeguards may add significant cost and resource requirements; additional complexity arises from the fact that to date in some European countries only aggregated data can be shared for secondary use and research purposes[78], and not pseudonymised data.

6.2 A final question that needs to be addressed, in light of the discussion so far conducted, is the following: is it still possible (and worth to bear the additional costs) to use full personal health data for research purposes?

The EHDS proposal does not repeal the provisions contained in the GDPR which regulate the further processing for research purposes of health data, therefore formally the answer should be pretty straightforward: nothing impedes the processing of personal health data for such purposes, as long as the GDPR regulation is complied with.

The real question is to whether there will be a concrete application of such model after the EHDS will be fully implemented. A key point to be highlighted concern the costs: while the GDPR model imposes organizational and compliance costs, the EHDS require data users to pay for a fee.

However, the pursuit of certain research projects may require the processing of personal data (complete with all their elements) and could not be carried out with anonymised or pseudonymised data[79]. In fact, undergoing such procedures depletes the quality of the data set: complete (and thus personal)

78 The borders of the concept have been clarified also by the EDPS in its preliminary opinion on data protection and scientific research", adopted on 6 January 2020 by the European Data Protection Board and the "EDPB Document on response to the request from the European Commission for clarification on the consistent application of the GDPR, focusing on health research", of 2 February 2021.

79 P. Quinn, *The Anonymisation of Research Data - A Pyric Victory for Privacy that Should Not be Pushed Too Hard by the EU Data Protection Framework?,* in European Journal of Health Law 24, 2017, pp. 1-21.

data allow integration with other datasets (e.g. health data previously collected under other circumstances, area of residence data, epidemiological data, etc.) and this provide research potentials related, for example, to the possibility of recontacting the data subjects for follow-up. This could only be done by tracing back the chain of purchase, although the data holder himself would not be able to maintain the access keys necessary for reidentification in the case of anonymization.

Differently, in the case of secondary processing of fully personal health data - whilst complying with data minimization and purpose limitation principles – it will be able to harness the full breadth and depth of the processed data. Thus, based on these premises, theoretically an interest in acquiring personal health data under the GDPR could remain. From a practical point of view, however, also in view of the rapid evolution of technology, which also by making use of artificial intelligence tools and by accessing large datasets makes it possible to obtain very precise analytics and sometimes even to derive elements that could only be gleaned from personal data, the benefits of acquiring health data through the EHDS mechanism could far outweigh those deriving from the (classical) processing of personal health data for research purposes.

7. Concluding remarks

7.1 The digital transformation in healthcare is based on data, and the EHDS proposal marks the latest and probably most comprehensive building block in the data strategy relating to health.

The citizen is thus given access to the data, in such a way to keep confidence and trust in the technological tools through which the processing is done. This creates more favourable conditions for use and re-use of health data, that produce both economical and societal benefits. In other terms, the proposal goal is both to enhance the patient's control over its data providing a full access to his medical records and granting portability rights and, at the same time, creating a single health data market.

As widely discussed above, the model for secondary use promoted by the EHDS is based on procedures aimed at limiting elements of 'personality' of data, still retaining their value for research. Therefore, to this extent anonymization and pseudonymization are essential. The notion of anonymized data – that in principle may be considered sufficiently clear – is however shaken by the application of advanced technologies[80] and its borders are *de facto* uncertain: indeed, the combination of (anonymized) big data by using tools with high analytical capabilities enhances greatly the capability for re-identification of anonymous data. Technologies enabling data mining and other phenomena that allow data to be inferred from those already available (often with high levels of accuracy) are capable of rendering established legal models ineffective. The technological advancements pose new challenges related e.g. to large-scale re-use and tying of personal data which seem difficult to reconcile with data protection principles such as purpose limitation and data minimisation. Hence, also from these findings it may arise the need for a refurbishment of the relevant rules. At the same time, in data-intensive research, anonymization would severely damage the potential for research as it would impede linking and updating data. Uncertainty still exists around what constitutes 'sufficient anonymisation' to transform personal data to non-personal data[81] and an effort towards standardization must be done.

7.2 How to streamline the procedure and thus make it easier to use personal health data, without neglecting the rights of the data subjects?

---

80 G. Comandè, *Research in the field of healthcare and data protection a puzzle ...resolvable*, Rivista Italiana di Medicina Legale (e del Diritto in campo sanitario), fasc.1, 1 february 2019, p. 199. See also in such regard S. Barocas – H. Nissenbaum, *Big data's end run around anonymity and consen*, in Privacy, big data, and the public good: Frameworks for Engagement. Cambridge University Press: Cambridge 2014, pp. 44-75; P. Quinn, *The Anonymisation of Research Data* cited; B.M. Knoppers et al., *Sampling populations of humans across the world: ELSI issues*, Annu Rev Genomics Hum Genet 2012; 13: 395-413; M. Mostert et al., *Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach*.

81 TEHDAS - Joint Action Towards the European Health Data Space, *Report on secondary use of health data through European case studies*, 28 February 2022.

The opting out option granted in the last version of the EHDS proposal provides the opportunity to the data subject to exempt their health data from this mechanism, without undermining the functioning of the system excessively (as would have been the case if an explicit opting-in mechanism had been envisaged). The need for express consent to opt-in to the sharing mechanism is, on the other hand, contemplated with reference to specific areas, which prove to be more sensitive. Thus, in these areas a more protective approach appears to be justified.

The idea of allowing data holders to collect a "broader consent"[82] which allows data controllers to carry out a number of diverse processing – some of them not precisely identified at the moment of the granting – and not strictly related to the primary use to be made of that data, is usually dismissed because appears to be, at least *prima facie*, incompatible with the guiding principles of the GDPR[83]. Although the GDPR does not directly contemplate it, one could envisage such a form of broad consent, provided it is accompanied by a process of oversight and approval of future research activities, as well as timely information to the data subject on the concrete use of the data and thus on the progress of the research[84]. As mentioned, there are some specific research fields in which the broad consent is already used – also because of the specific characteristics of the data employed – like genomic research[85]. Some cautions must me expressed however with reference to the risk of dissemination of those data which could cause very serious damage to the data subject, his social

82 M. A. Rothstein et al., *Broad Consent for Future Research*, Ethics & Human Research, November-December 2018, Vol. 40, No. 6 pp. 7-12, available at https://www.jstor.org/stable/10.2307/26776205.

83 With regard to the specificity of the research uses for which consent is given, ECDP guidelines on consent (no. 5/2020, adopted on 4 May 2020) identify examples "*allowing data subjects to consent to a research purpose in more general terms and to specific phases of a research project that it is known from the outset will take place. As the research progresses, it will then be possible to obtain consent for later phases of the project before the corresponding phase begins […] still be in line with the deontological standards applicable to scientific research*". See p. 158 of the mentioned guidelines.

84 Similarly, C. Grady et al., *Broad Consent For Research With Biological Samples: Workshop Conclusions*, Am J Bioeth. 2015; 15(9), p. 6, available at doi: 10.1080/15265161.2015.1062162.

85 D. Hallinan, *Broad consent under the GDPR: an optimistic perspective on a bright future*, Life Sci Soc Policy. 2020 Dec; 16: 1, available at doi: 10.1186/s40504-019-0096-3.

and working life. In such a case, the potential stigma arising from the dissemination of this data as a result of its mishandling could be so serious that it could not counterbalance the procedural simplification.

Moreover, the innovations contained in the EHDS proposal overcome many of the practical challenges that would suggest the adoption of forms of broad consent (while at the same time encouraging the use of pseudonymised or anonymised data, in deference to the principle of minimisation).

Still the EHDS proposal appears to have troublesome and challenging profiles, which would require rethinking or at least more careful consideration. For example, it seems to provide an insufficient control to patients over the sharing and uses of their personal data, granting them a mere power to power to deny their use outright. Although claiming that it would provide more control to individuals over their private information, the EHDS proposal appears to deprive them of this control, specifically when it comes to secondary use. In fact, patients would be granted a mere right of opting out from the secondary use of their data (generally considered), having no say over the specific sharing and commercial exploitation of their data, which will be decided solely by the data holders with the consent of the competent public bodies. In addition, even when data subjects have exercised their right to opt out, Member States have the power to adopt laws aimed at overriding that choice, in order to allow secondary use by entities in charge of public tasks in the field of public health, in specific situations in which there is a public interest. This provision, which is clearly expressive of the relevance of public health interests, might however threaten to (excessively) sideline the rights of the data subject, unless it is guaranteed that the personal elements of the data are removed already at the time of their storage of the data (and not only) when shared for re-use.

7.3 As for the many strongly positive profiles, in the first place, the EHDS proposal has the merit of providing a framework capable of favouring

the secondary use of health data, thus promoting the scientific progress and its positive effects for public health. This can be grasped, for example, from the provision which mandate those who have had access to the data (and conducted research on it) to publish the relevant results. In this way, it is ensured that also private research may bring public health benefits[86].

Moreover, the EHDS proposal must be commended for envisaging an effective and structured health data sharing model that in absence of specific regulatory obligations or incentives would hardly happen, also because of the costs to incur for setting up the data collection (and storage) and making their systems and data interoperable. Mandatory sharing of data (and the relevant tariff mechanisms provided for access to health data for secondary use) as well as the set-up of common standards may therefore represent appropriate stimuli for the establishment of functional and efficient practices. Although the cost of the initial investment required from data holders is not fully passed on to potential buyers, the mechanism of health data trading (for a consideration) makes the system more viable, it is still unlikely that it will get off the ground without substantial public investment[87].

The EHDS secondary use mechanism has the ultimate goal of creating a functioning 'network' system capable of yielding benefits to patients and well-being to the community.

The health data contained in hospital records will thus finally be considered not only as a summary of informative elements relating to the medical or diagnostic procedures performed, but as an asset to be valorised in view of its possible further uses: medical (such as the development of effective treatments based on the analysis of the results of clinical trials on a sample of patients),

86 See Article 46 (11) EHDS Proposal.

87 Article 32a shows awareness of the costly implications of compliance with the EHDS mechanisms and, applying the principle of proportionality, provides that individual researchers and micro-businesses are exempted from the sharing obligation of the health data in their possession.

diagnostic, research and even commercial, within the boundaries of permitted uses as deemed worthy.

Lastly, it still remains the challenge of the effective implementation of the EHDS, whereby member States will have to ensure that all entities operate using common standards and will actually upload data to the appropriate databases. Otherwise, the functioning of the entire mechanism will be compromised with negative consequences both for local communities and for the progress of European research. The economic incentives deriving from secondary use fees are likely to be insufficient if not supported by a determined political effort and cultural revolution.