European Journal of Risk Regulation (2024), **15**, 928–938
doi:10.1017/err.2024.67

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Anonymisation, Pseudonymisation and Secure Processing Environments Relating to the Secondary Use of Electronic Health Data in the European Health Data Space (EHDS)

Richard Rak

University of Milan, Department of Legal Sciences "Cesare Beccaria", Information Society Law Center, Italy
Email: richard.rak@guest.unimi.it

Abstract

By establishing a common data governance mechanism across the EU, the Regulation on the European Health Data Space (EHDS) aims to enhance the reuse of electronic health data for secondary use (e.g. public health, policy-making, scientific research) purposes and realise associated benefits. However, the EHDS requires health data holders to make available vast amount of personal and non-personal electronic health data, including electronic health data subject to intellectual property (IP) rights, for secondary use, which may pose risks for stakeholders (patients, healthcare providers and manufacturers alike). This paper highlights some conceptual legal problems which need to be addressed in order to provide clearer regulatory requirements to ensure effective and consistent implementation of key data minimisation measures (anonymisation or pseudonymisation) and data management safeguards (secure processing environments). The paper concludes that the EHDS has been drafted ambiguously (for example, its definition of "electronic health data" or the list of "minimum categories of electronic data for secondary use"), which could lead to inconsistent data management practices and may impair the rights and legitimate interests of data subjects and rights holders. To address legal uncertainties, prevent fragmentation and mitigate/eliminate risks, the EHDS requires closely coordinated implementation and legislative fine-tuning.

Keywords: European Health Data Space (EHDS); data protection; electronic health data

I. Introduction

The European Health Data Space (EHDS) provides common EU-wide rules in order to facilitate the digital processing of electronic health data for primary and secondary use purposes through secure and interoperable means.[1] While "primary use" refers to the processing of personal electronic health data for the provision of healthcare to the natural person to whom that data relates, "secondary use" refers to the reuse of personal and non-personal electronic health data for purposes other than the initial purposes for which they were collected or produced (such as for public health, policy-making and regulatory activities, official statistics, or scientific research including innovation activities). This

[1] Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space (henceforth: "EHDS"), compromise text. Available: https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=celex:52022PC0197; https://www.consilium.europa.eu/en/press/press-releases/2024/03/15/european-health-data-space-council-and-parliament-strike-provisional-deal/. Note that the final adopted and published version of the EHDS may slightly differ from the compromise text analysed in this paper.

paper focuses on the common data governance mechanism established by the EHDS that sets rules for the secondary use of electronic health data in the EU. The scaling up of the secondary use of electronic health data is a broadly supported policy objective in the European health ecosystem. According to an estimate, 97% of data collected or produced by hospitals goes unused, leaving many potentially life-saving medical insights or efficiency gains locked away.[2] However, the realisation of potential benefits from permitting the reuse of vast amount of electronic health data requires clear regulatory requirements and the implementation of data processing and other data management safeguards in a responsible, risk-aversive, trustworthy and consistent manner.

The common data governance mechanism for secondary use of electronic health data in the EHDS requires health data holders to make available datasets that fall under specific data categories (e.g. data from electronic health records, data from clinical trials, data from medical devices) for secondary use purposes. This also entails an obligation to make available electronic health data covered by intellectual property (IP) rights, trade secrets and/or regulatory data protection. A description of those datasets shall be made public in national and EU datasets catalogues. The governance of the common data governance mechanism is overseen by health data access bodies (HDABs) acting as public authorities. Each Member State designates one or more HDABs responsible for overseeing access to electronic health data. To enable cross-border access to electronic health data for secondary use, Member States and the Commission have agreed to operate the HealthData@EU infrastructure connecting national contact points (organisational and technical gateways). Overall, this data governance framework can enable a health data applicant to submit a data access application to the competent HDAB (or a trusted health data holder) in order to request access to electronic health data. If the health data applicant can demonstrate a justified purpose and adequate guarantees, the competent HDAB may issue a data permit that allows the health data applicant to access electronic health data as a health data user in a secure processing environment, in principle, in an anonymised format or, if necessary, in a pseudonymised format. Alternatively, a health data applicant may submit a health data request with the aim of obtaining an answer only in an anonymised statistical format from a HDAB.

Given that the EHDS introduces several new (but ambiguously defined) concepts and an ambitious (but untested) common data governance mechanism for secondary use of electronic health data, it warrants analyses of what legal challenges and risks may stakeholders face. The paper sheds light on some of the most important legal challenges that need to be addressed to ensure clearer regulatory requirements and to facilitate effective implementation of key data minimisation measures (anonymisation or pseudonymisation) and other data management safeguards (secure processing environments) within this framework. The analysis focuses on how the legislative phrasing of the EHDS may pose practical challenges in implementation and proposes alternative regulatory solutions for consideration.

II. Determining the scope of electronic health data for secondary use under the EHDS and the relevance of "pseudonymisation" and "anonymisation"

1. The scope of "personal electronic health data" under the EHDS

One of the basic legal questions when determining the scope of electronic health data for secondary use is how the EHDS defines the contours of "electronic health data" and how it

[2] Judith Moore, Yasmin Dias Guichot, "How to harness the power of health data to improve patient outcomes," World Economic Forum (5 January 2024). Available: <https://www.weforum.org/agenda/2024/01/how-to-harness-health-data-to-improve-patient-outcomes-wef24/>.

delineates its subsets. Regarding the definition of "personal electronic health data" (as one of the subsets of "electronic health data"), the EHDS provides that it should encompass "data concerning health" and "genetic data," as defined under Articles 4(13) and 4(15) of Regulation (EU) 2016/679 ("GDPR")[3], that are processed in electronic form. The cross-references to the two data categories under the GDPR (and their interpretations under EU data protection law) may shape the understanding of the notion of "personal electronic health data" in the following ways.

With reference to Article 4(15) of the GDPR, the notion of "personal electronic health data" under the EHDS would cover personal data processed in an electronic form that is "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." Recital (35) of the GDPR provides further explanation by adding that "[p]ersonal data concerning health should include all data pertaining to the health status of a data subject which reveal information relating to the past, current or future physical or mental health status of the data subject." However, the combined reading of Article 4(15) and Recital 35 of the GDPR may lead to a dubious interpretation. A question may arise, notably whether: (a) only data that are already related to the health status of a natural person *and* are revealing of information about their health status are protected as special categories of personal data [if the emphasis of the interpretation is on Article 4(15)]; or (b) any data revealing information about the health status of a natural person are protected as special categories of personal data [if the emphasis of the interpretation is on the elaboration in Recital 35].[4] In the absence of a binding interpretation, it is relevant to recall that the Article 29 Data Protection Working Party (the predecessor of the European Data Protection Board) explained that "the term "data revealing . . . " is to be understood that not only data which by its nature contains sensitive information is covered [.], but also data from which sensitive information with regard to an individual can be concluded."[5] Although the regulatory landscape has changed (after the GDPR repealed Directive 95/46/EC), the interpretation of the Article 29 Data Protection Working Party may remain indicative on when personal data constitutes personal (electronic) health data (in a digital context). Accordingly, this is the case when:[6]

1. data about the physical or mental health status of a data subject is generated in a professional medical context, including data related to contacts with patients and their diagnosis (which does not necessarily imply "ill health") and treatment by providers of health services, data on diseases, disabilities, medical history and clinical treatment, as well as data generated [or obtained] by devices or apps about a data subject, which are used in this context, irrespective of whether they qualify as "medical devices";

[3] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (henceforth: "GDPR").

[4] Gianclaudio Malgieri, Giovanni Comandé, "Sensitive-by-distance: quasi-health data in the algorithmic era" (2017) 26(3) Information & Communications Technology Law 229, 232. DOI: <https://doi.org/10.1080/13600834.2017.1335468>.

[5] Article 29 Data Protection Working Party, Advice paper on special categories of data ("sensitive data") (4 February 2011), 6. Available: <https://ec.europa.eu/justice/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf>.

[6] Article 29 Data Protection Working Party, Letter from the ART 29 WP to the European Commission, DG CONNECT on mHealth, Annex – health data in apps and devices (5 February 2015), 3. Available: <https://ec.europa.eu/justice/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf>.

2. there is a "demonstrable relationship" between the data and the capacity to determine a health aspect (health status or health risk) of a person based on the data itself or on the data in combination with data from other sources;
3. conclusions (such as inferences) are drawn about a person's health status or health risk (irrespective of whether those conclusions are accurate or inaccurate, legitimate or illegitimate, or otherwise adequate or inadequate).

The second element of the definition of "personal electronic health data" is a cross-reference to Article 4(13) of the GDPR. With regard to that provision, the notion of "personal electronic health data" under the EHDS would also cover genetic data processed in an electronic form "relating to the inherited or acquired genetic characteristics of a natural person which give unique information about the physiology or the health of that natural person and which result, in particular, from an analysis of a biological sample from the natural person in question." It is important to emphasise that the GDPR refers "in particular" to a sample collected directly from the donor. That implies that genetic data obtained directly from donors are personal data concerning them, and that genetic data relating to a natural person obtained by other means could also be considered as such. However, in the case of a biological sample, it is debatable whether it is *per se* genetic data. An affirmative argument could be that a sample contains data (e.g. DNA) and merits the protection of data protection law.[7] A counter-argument to this could be that a sample is merely a physical matter, which can only be the source of the data that is explicitly defined by the GDPR.[8] Another categorisation problem is that the notion of "genetic data" is too narrow to capture all "omics data" referred to by the EHDS. The only way non-genetic omics data may qualify as "personal electronic health data" under the EHDS is if they fall within the scope of "data concerning health" under the GDPR. Finally, it is important to point out that while genetic information is unique and distinguishes an individual from other individuals, it may at the same time reveal information about and have implications for that individual's blood relatives (biological family), and may even characterise a broader group of people (e.g. ethnic communities, patients with a similar disease).[9] What follows from this is that biological family members could be considered data subjects along with the original donor when their common genetic data is processed. Even genetic data concerning deceased persons or unborn children (foetuses) could be considered as personal data concerning their (living) biological family members.[10] These overlaps would require legal clarification, as an overly broad interpretation of the notion of "genetic data" could lead to legal uncertainties in the EHDS regarding the application of the rights of natural persons in relation to the processing of genetic data concerning them as "personal electronic health data" for primary and secondary use purposes.

[7] Dara Hallinan, Paul De Hert, "Many Have It Wrong – Samples Do Contain Personal Data: The Data Protection Regulation as a Superior Framework to Protect Donor Interests in Biobanking and Genomic Research" in Brent Daniel Mittelstadt, Luciano Floridi (eds), *The Ethics of Biomedical Big Data. Law, Governance and Technology Series*, vol. 29 (Springer, 2016), 119. DOI: <https://doi.org/10.1007/978-3-319-33525-4_6>.

[8] Mahsa Shabani, Pascal Borry, "Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation" (2018) 26 European Journal of Human Genetics 149, 152. DOI: <https://doi.org/10.1038/s41431-017-0045-7>.

[9] Article 29 Data Protection Working Party, Working Document on Genetic Data (14 March 2004), 4. Available: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2004/wp91_en.pdf>.

[10] Taner Kuru, Iñigo de Miguel Beriain, "Your genetic data is my genetic data: Unveiling another enforcement issue of the GDPR" (2022) 47 Computer Law & Security Review 105752, 4. DOI: <https://doi.org/10.1016/j.clsr.2022.105752>.

2. The scope of "electronic health data" under the EHDS

Another legal question that may fundamentally shape the implementation of the rules on the secondary use of electronic health data stems from the problem of how the EHDS draws the boundary of the scope of data that falls under the larger scope of "electronic health data" without qualifying as "personal electronic health data." Following significant amendments in the course of the legislative procedure, the compromise text of the EHDS defined this category as "non-personal electronic health data" meaning "electronic health data other than personal electronic health data, encompassing both data that has been anonymised so that it no longer relates to an identified or identifiable natural person and data that has never related to a data subject". One of the major problems with this definition is that it does not clarify what is its link with the health status of natural persons or other healthcare-related information. Furthermore, it generates uncertainty that the cases covered by the phrase "data that has never related to a data subject" relies on an assumption that there is some kind of clear threshold between non-personal data and non-personal electronic health data. However, this is not the case. For example, connected products in healthcare or care sectors (e.g. connected medical devices, wellness applications, sensors in ambient assisted living systems, or certain product components of hospital information systems or healthcare platforms) and related services (e.g. algorithms/software or AI systems enabling the functioning of a connected product) may generate or obtain vast amount of data (e.g. data relating to hardware status, battery levels, malfunctions, data transmissions, version control, security functions or the location of the product) where there is arguably no clear "demonstrable relationship" between the data and the capacity to determine the health aspect of a natural person. The legal requirement under the EHDS for health data holders to make available all such data for secondary use purposes may pose security risks for all parties concerned (including data subjects, health data holders, rights holders, manufacturers and other economic operators) and increase compliance, data storage and associated environmental costs. It is also unclear how the requirement to make available some of these datasets under the EHDS would interact with corresponding (and arguably conflicting) requirements under the Data Act.

With regard to the uncertainties and potential risks affecting the secondary use of electronic health data in the EHDS, it could be useful to highlight alternative ways in which the co-legislators could have regulated (or could still fine-tune) the definition of "electronic health data." One solution would be to significantly narrow the scope of electronic health data made available under the EHDS for secondary use to cover only "personal electronic health data." In essence, this would mean narrowing the definition of "electronic health data" to "personal electronic health data." Another solution would be to make a minor edit to the current definition by clarifying that non-personal electronic health data should relate to the health status or genetic characteristics of unidentified or unidentifiable natural persons. Finally, a more structural revision of the definition of "electronic health data" (which was suggested by industrial actors during the legislative procedure) would cover the following data categories:[11]

1. "personal electronic health data," including "personal electronic health data in pseudonymised format";
2. "anonymised electronic health data"; and
3. "anonymous statistical electronic health data".

The latter solution would have arguably improved legal consistency with the GDPR and within the EHDS by facilitating a more consistent interpretation of the concepts of

[11] DIGITALEUROPE, "European Health Data Space (EHDS): key issues to address in trilogues" (22 December 2023), 3–4. Available: <https://cdn.digitaleurope.org/uploads/2024/01/EHDS-trilogues-DIGITALEUROPE-position-paper-1.pdf>.

"anonymisation" and "pseudonymisation." In connection with the notion of "personal electronic health data," it would have been useful to clarify in the EHDS (practically, among its Recitals) a set of criteria for data to qualify as "personal electronic health data" by considering the Article 29 Data Protection Working Party's abovementioned opinions. As a supplementary definition, it could have been useful to define "personal electronic health data in pseudonymised format" as personal electronic health data that has undergone pseudonymisation, in accordance with Article 4(5) of the GDPR. As such data can be attributed to a natural person by the use of additional information, it should be considered to be information on an identifiable natural person, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal electronic health data are not attributed to an identified or identifiable natural person. However, considering that pseudonymisation reduces the linkability of a dataset with the original identification of a data subject, it would bring further clarity if the EHDS explicitly stated that those datasets have distinct characteristics. Regarding the other proposed categories of data, "anonymised electronic health data" could have been defined as "data obtained as the result of processing (anonymising) personal electronic health data in such a manner that natural persons are not identifiable and cannot be re-identified by any means reasonably likely to be used by the [electronic] health data holder or to whom the data is made available, in particular the [electronic] health data user." "Anonymous statistical electronic health data" could have been defined as "data obtained as the result of any operation of data collection, processed in electronic form, that is related to the health status, health risk or genetic characteristics of natural persons for statistical purposes in such a manner that natural persons are not identifiable and cannot be re-identified by any means reasonably likely to be used by the [electronic] health data holder or to whom the data is made available, in particular the [electronic] health data user."

The definitions of these two latter categories of data would have improved legal clarity. Firstly, they incorporate the "reasonableness" test. With regard to Recital (26) of the GDPR and the Article 29 Data Protection Working Party's Opinion 05/2014 on Anonymisation Techniques,[12] the "means reasonably likely to be used" is the criterion to assess whether an anonymisation process is sufficiently robust, i.e., whether identification has become "reasonably" impossible. In the context of the EHDS, that test could be applicable (relatively) from the perspective of the health data holder or any other person (in particular, the health data user) to whom anonymised electronic health data or anonymous statistical electronic health data are made available. Secondly, the proposed definitions would overcome the abovementioned problem of having to determine a threshold where "non-personal data" may become "related to health." Thirdly, the addition of the notion of "anonymous statistical electronic health data" would ensure that data collections for statistical purposes (official or not) that are related to the health aspects or genetic characteristics of natural persons are also covered by the definition of "electronic health data." It would provide a sufficient degree of clarity that this category of data would be defined by the intended purpose and methodology of the data operation (i.e. "statistical purposes related to the health status, health risk or genetic characteristics of natural persons"). The introduction of the notion of "anonymous statistical electronic health data" would also ensure coherence and clarity to the provisions of the EHDS that require that an answer to a health data request should be provided by HDABs in an anonymised statistical format.

[12] Article 29 Data Protection Working Party, Opinion 05/2014 on Anonymisation Techniques (10 April 2014), 8. Available: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>.

3. The requirement for health data holders to make available certain categories of (electronic health) data for secondary use

The EHDS enumerates the "minimum categories of electronic data" that health data holders shall make available for secondary use. However, uncertainties persist about the exact scope of those data categories. In this regard, a multi-stakeholder coalition of 36 European patient organisations, medical-professional organisations, health research infrastructures and industry associations called for the legislative refinement of the EHDS before the EU interinstitutional negotations arguing that the EHDS should clearly specify the scope of electronic health data categories for secondary use. The informal stakeholder group explained that: "[t]he scaling up of the secondary use of electronic health data under a harmonised data governance framework could bring wide-ranging benefits to healthcare-related activities and research in the EU, if the associated risks are eliminated or sufficiently mitigated. It is also important to ensure consistency in the use of terminology, as it would lead to uncertainty if the various data categories apply to "data," "aggregated data," "electronic data," "health data," "healthcare-related data," "determinants of health" or "electronic health data," without there being any clear indication about what some of these data categories would entail. It would also cause uncertainty if certain data types were to fall into multiple data categories, but specific provisions would add particular conditions (e.g. aggregated form, opt-out mechanism) to make them available for secondary use."[13] Despite the joint call from stakeholders in the health ecosystem, most of these concerns were not addressed by the co-legislators in the final legislative phase of the EHDS.

Before providing access to electronic health data for secondary use purposes, the EHDS requires the implementation of anonymisation or pseudonymisation techniques. However, the effective implementation of those safeguards and accompanying measures depends on a case-by-case basis. In the case of certain electronic health data categories, their implementation may cause considerable challenges. The most often mentioned example is the problem of anonymising human genetic (or other omics) data or personal data from biobanks due to the possible risk of re-identifying data subjects. Although the linkage of multiple datasets and the potential insights drawn therefrom in a big data context may enable the realisation of benefits to advance personalised healthcare and health research and innovation, it may also increase the risks of re-identification of data subjects and may generate additional privacy risks. To mitigate those risks and provide legal assurances, it would be crucial that the EHDS specifies the electronic health data categories for secondary use by clearly defining the registries/databases from which electronic health data shall be made available and/or the original processing purposes of (personal) electronic health data. This could be clarified in one of the implementing acts of the EHDS that aims to "set out the minimum elements health data holders are to provide for datasets and their characteristics." Furthermore, there is a need to provide clarity and consistent interpretation on anonymisation and pseudonymisation techniques in the new legal landscape. This matter could be addressed by the EDPB by issuing a guideline on the processing of personal data for scientific research purposes and/or specifically on the topics of anonymisation and pseudonymisation.[14]

In addition to data protection challenges, the EHDS poses risks to the effective protection of IP, as it requires electronic health data protected by IP rights, trade secrets

[13] Stakeholder coalition calls for legislative refinement of the EHDS (4 December 2023). Available: <https://www.digitaleurope.org/news/stakeholder-coalition-calls-for-legislative-refinement-of-the-ehds>.

[14] European Data Protection Board, EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research (2 February 2021), 11. Available: <https://edpb.europa.eu/sites/default/files/files/file1/edpb_replyec_questionnaireresearch_final.pdf>.

and/or covered by the regulatory data protection right to be made available for secondary use. In such cases, "health data holders shall inform the health data access body of and identify any electronic health data containing content or information protected by intellectual property rights, or trade secrets and/or covered by the regulatory data protection right" and the "health data access bodies shall take all specific appropriate and proportionate measures, including legal, organisational, and technical ones, they deem necessary to preserve the protection of intellectual property rights, trade secrets and/or the regulatory data protection right." However, this radically new IP governance scheme has been criticised for failing to provide adequate and effective control and safeguards to health data holders (or other rights holders) and may undermine existing legal protection and incentives (provided under international, EU and national laws) that are vital for researchers and innovators.[15]

III. Data minimisation and purpose limitation requirements in the governance of the secondary use of electronic health data under the EHDS

The EHDS requires HDABs to provide access to electronic health data in accordance with the principles of data minimisation and purpose limitation. In connection with that requirement, the EHDS makes reference to the corresponding data protection principles under Article 5(1) of the GDPR. However, the data protection principles under the GDPR are only applicable in the context of processing personal data. Considering that the EHDS would require "electronic health data" to be made available for secondary use (i.e. not only "personal electronic health data"), it would extend the application of data protection principles to operations performed on all "electronic health data" (including non-personal electronic health data). It is one of several problematic examples where the EHDS blurs the borderline between the protection of "personal electronic health data" and "non-personal electronic health data." This inconsistency generates legal uncertainty and poses risks.

According to the data minimisation and purpose limitation requirements set forth under the EHDS, the "health data access body shall ensure that access is only provided to requested electronic health data that are adequate, relevant and limited to what is necessary in relation to the purpose of processing indicated in the data permit application by the health data user and in line with the data permit granted." In terms of the format in which the health data access body shall provide access to electronic health data, the EHDS provides for two ways. In principle, the HDAB shall provide the electronic health data in an anonymised (or anonymised statistical) format, where the purpose of processing by the health data user can be achieved with such data, taking into account the information provided by the health data user. If the health data user demonstrates that its processing cannot be achieved with anonymised data, taking into account the information provided by the health data user, the HDAB shall provide access to (personal) electronic health data in pseudonymised format, where the information necessary to reverse the pseudonymisation shall be available only to the HDAB (or a body that acts as a trusted third party in accordance with national law).

However, the aforementioned rules have certain shortcomings. In the principal case, it is unclear whether "anonymised (or anonymised statistical) format" should encompass electronic health data that was collected anonymously from the outset, or only data obtained as the result of anonymising personal electronic health data. This exemplifies that the inconsistent use of the wording "anonymous" or "anonymised" electronic health data may lead to different legal interpretations. In addition to this problem, there is also a lack of clarity about whether the EHDS requires "absolute" or "relative" anonymity. Absolute anonymity means that re-identification is impossible for any third party,

[15] DIGITALEUROPE (n 11), 8–10.

whereas relative anonymity means that re-identification is reasonably unlikely from the perspective of the controller in relation to the circumstances of the case (taking into account objective factors, such as the possible costs of and the amount of time required for identification, as well as technological developments, and subjective factors, such as the specific capacities of relevant actors). In general, European data protection supervisory authorities and the case law of the Court of Justice of the European Union (CJEU)[16] lean towards favouring a relative understanding, or embrace a half-way test between the absolute and relative approaches.[17] In the absence of a generally applicable guidance on anonymisation, it would have been important to clarify this matter in the EHDS, for example, by clearly defining the subsets of "electronic health data," as suggested above.

The allocation of data protection (and other data management) responsibilities relating to the secondary use of (personal) electronic health data is another important matter for consideration under the EHDS. Theoretically, the duty of performing anonymisation and pseudonymisation tasks could be allocated to the health data holder, the HDAB or a separate dedicated body. Although there were conflicting positions during the legislative procedure, the EHDS eventually provides that pseudonymisation and anonymisation can be carried out by the HDAB or by the health data holder (or their processors if they are acting in the capacity of a controller) as early as possible in the chain of making data available for secondary use. Indeed, if the health data holder is able to perform those tasks, then it may provide additional security reassurances, as the health data holder may often be in a more suitable position (than the HDAB) to perform those tasks with regard to the circumstances of the original processing and related data protection factors.

IV. Secure processing environments in the EHDS and considerations for their implementation

1. The purpose of secure processing environments

In accordance with the data minimisation requirement under the EHDS, the HDAB (or a trusted health data holder) must provide access to electronic health data pursuant to a data permit in a secure processing environment, which complies with technical and organisational measures as well as security and interoperability requirements. A "secure processing environment" is defined under Article 2(20) of the Data Governance Act as "the physical or virtual environment and organisational means to ensure compliance with Union law, such as Regulation (EU) 2016/679, in particular with regard to data subjects' rights, intellectual property rights, and commercial and statistical confidentiality, integrity and accessibility, as well as with applicable national law, and to allow the entity providing the secure processing environment to determine and supervise all data processing actions, including the display, storage, download and export of data and the calculation of derivative data through computational algorithms." Some EU Member States have already established or recognised SPEs (also known as "trusted research environments" (TREs)). However, there are no common international, EU or domain-specific standardisation and certification requirements for SPEs. For this reason, there is a need to set a roadmap for the harmonised development of SPEs in the EU.[18] This could

[16] *Case C-582/14, Patrick Breyer v Bundesrepublik Deutschland [2016] ECLI:EU:C:2016:779*; *Case T-557/20, Single Resolution Board v European Data Protection Supervisor [2023] ECLI:EU:T:2023:219*.

[17] Michèle Finck, Frank Pallas, "They who must not be identified—distinguishing personal from non-personal data under the GDPR2 (2020) 10(1) International Data Privacy Law 11, 17. DOI: <https://doi.org/10.1093/idpl/ipz026>.

[18] Irene Schlünder, Michaela Th. Mayrhofer, Erdina Ene, "Elements of Secure Processing Environments" (HealthyCloud and EOSC-Life Workshop Report v1.0, Brussels/online, 19–20 June 2023), 12. DOI: <https://doi.org/10.5281/zenodo.8341642>.

allow for a plurality of SPE providers that, in return, share common building blocks (e.g. standard operating procedures) to ensure interoperability. It is also important to recognise that there is no one-fits-all solution for SPEs, as technology changes rapidly and scientific research needs flexibility. Due to this consideration, data processing requirements in SPEs need to be balanced with accessibility and user needs. Overall, as such developments need to be user-driven, bottom-up approaches may be more effective than top-down solutions in this regard. This is important to consider for the implementing act of the EHDS that is intended to set forth "the technical, organisational, information security, confidentiality, data protection and interoperability requirements for the secure processing environments."

2. *Requirements for secure processing environments in the EHDS*

Although the EHDS does not make it explicitly clear in the binding part of the text, secure processing environments (SPEs) could be operated by a HDAB, a trusted health data holder (such as a large hospital or a research infrastructure with adequate resources), or a third party SPE provider (operating under the supervision of a HDAB or a trusted health data holder). While there are arguably benefits in allowing trusted health data holders to provide access to electronic health data pursuant to a data permit in a SPE, a potential legal problem is that trusted health data holders have to fulfil the tasks of HDABs. In other words, they are required to act *de facto* as HDABs, despite not having that legally enshrined capacity.

Regarding the rules on the functionalities of SPEs, the EHDS requires HDABs (or trusted health data holders) to ensure that electronic health data can be uploaded by the health data holder (in the format determined by the data permit) and can be accessed by the health data user in a SPE. Considering that there may be interoperability challenges affecting the uploading of electronic health data in a given format, this issue may require collaboration between the health data holder, the HDAB and, where relevant, the trusted health data holder and/or third party SPE service provider. On the other side, the EHDS states that health data users may only download non-personal electronic health data, including electronic health data in an anonymised statistical format, from the SPE. This may cause practical challenges. For example, if personal electronic health data are made available in the SPE in a pseudonymised format, then that would not be downloadable, regardless of the fact that the health data user would already have access to it and may have performed operations (data analysis) on it. It is also questionable how this download restriction could affect the development and further deployment of algorithms for AI systems in healthcare, which may often need to be trained, validated and tested on personal electronic health data in a pseudonymised format to meet safety and performance requirements.

V. Conclusions

The legislative procedure of the EHDS has highlighted the complexities of establishing an EU-wide common data governance mechanism for the secondary use of electronic health data. Although the EHDS requires the integration of necessary safeguards (including anonymisation, pseudonymisation and secure processing environments), their implementation is intertwined with several factors in the EHDS. This paper explains that certain provisions of the EHDS are ambiguous, which could lead to inconsistent data management practices and may impair the rights and legitimate interests of data subjects and rights holders. For example, the lack of clarity that stems from the broad definition of "electronic health data", especially the ambiguous scope of "non-personal electronic health data",

may amplify risks when such data is required to be made available for secondary use from connected devices in healthcare. Another example concerns the challenge of anonymising genetic (and other omics) data and the potential risks of revealing sensitive information about other individuals belonging to the data subject's biological family. Significant uncertainties and risks also persist concerning the scope of electronic health data categories that health data holders shall make available for secondary use in the EHDS. There is also a lack of effective control granted to rights holders when electronic health data subject to their IP rights are made available for secondary use. Overall, with regard to the magnitude and number of problems presented in this paper, it is essential that the Commission and Member States coordinate to address legal uncertainties, prevent fragmentation and mitigate/eliminate risks in order to avoid any detrimental consequences in the implementation of the EHDS. As digital health is a multidisciplinary and fast-evolving field, the engagement of the health ecosystem could help to fine-tune critical points in the EHDS and ensure that the adoption of relevant implementing acts can facilitate the realisation of the intended policy goals.

Competing interests. The author has no conflicts of interest to declare.
Disclaimer: The opinion expressed in this paper are those of the author only and should not be attributed to any organisation, unless indicated.

Cite this article: R Rak (2024). "Anonymisation, Pseudonymisation and Secure Processing Environments Relating to the Secondary Use of Electronic Health Data in the European Health Data Space (EHDS)". *European Journal of Risk Regulation* **15**, 928–938. https://doi.org/10.1017/err.2024.67