Legal bases for effective secondary use of health and genetic data in the EU: time for new legislative solutions to better harmonize data for cross-border sharing?

Regina Becker *, Davit Chokoshvili**, Edward S. Dove ***

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

Key points

- The secondary use of health and genetic data between different actors and countries can be medically and scientifically rewarding, but it requires a solid legal framework to enable responsible cross-border and cross-sector access to such data.
- From an EU data protection law standpoint, such secondary use requires a legal basis under Article 6(1) and a permission under Article 9(2) of the EU General Data Protection Regulation (GDPR). However, we argue that, in practice, a suitable legal basis is not always available to all stakeholders to process health and genetic data for secondary uses.
- The EU has recently proposed a European Health Data Space (EHDS), which is the first common EU 'data space' in a specific area to emerge from the 'European strategy for data'. Among other aims, the EHDS seeks to provide a consistent, trustworthy, and efficient system for reusing health and genetic data for research, innovation, policy-making, and regulatory activities (ie, secondary use). Yet, the EHDS does not account for the different phases in the data reuse lifecycle in a sufficiently encompassing manner, resulting in missing or insufficient GDPR legal bases for undertaking certain crucial processing operations.
- As a result of the dependency on legislative acts and the varying nature of stakeholders involved, we find that there can be significant hurdles for secondary use, and some actors can even be excluded from participation entirely. Consequently, we advocate new data protection legislative solutions to harmonize data for cross-border sharing.

Introduction

In today's data-driven world, organizations rely on growing amounts of personal data. Health and genetic data are no exception, and indeed these are often the most valued personal data of all, both by individuals who view them as, generally speaking, more sensitive than other kinds of data, and organizations, which consider them to have greater monetary and innovation value than many other kinds of data. Those collecting and using these data include healthcare providers such as hospitals and clinics, as well as biomedical research organizations that collect rich medical data from patients and/or research participants.[1] These organizations typically do so in the pursuit of their mission, such as administering health care to a patient, or undertaking a particular research project.

However, there is a growing recognition that such data, collected for a *specific primary* use, holds significant utility for *future, potentially unspecified* uses, and

* Regina Becker, Luxembourg National Data Service, Luxembourg
** Davit Chokoshvili, Luxembourg National Data Service, Luxembourg
*** Edward S. Dove, School of Law and Criminology, Maynooth University, Ireland. Email: edward.dove@mu.ie.
The authors thank Ilaria Colussi, Giovanni Comandé, Fruzsina Molnár-Gábor, Irith Kist, Guillermo Lazcoz, Pilar Nicolás Jiménez, Susanne Rebers, Marjanka Schmidt, Adrian Thorogood, Olga Tzortzatou-Nanopoulou, Bert Verdonck, Alexandra Ziaka, as well as the Ethical, Legal and Social Implications (ELSI) working group of the 1+ Million Genome Initiative, for their helpful comments and suggestions.

1 Ivo D Dinov, 'Volume and Value of Big Healthcare Data' (2016) 4 Journal of Medical Statistics and Informatics 3.

© The Author(s) 2024. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

for purposes *unrelated* to the initial collection. This is often described as 'secondary use' of the data.[2] As noted elsewhere, there is no broad consensus as to the precise definition of 'secondary use' of data, but for the purposes of this article, we broadly refer to 'secondary use' of personal health data in a manner that encompasses any use of the data beyond the purpose for which the data were initially collected or generated.[3]

There are various ways in which secondary uses may arise. For example, biomedical research organizations may re-utilize existing data internally for additional research projects addressing new questions, or share them with their close collaborators, for example, within a research consortium. However, the most meaningful societal benefits stem from making already-collected data systematically available to third parties in a streamlined and scalable way, who can then use these data for their own purposes.[4] Disease registries are well-known examples for supporting the quality and outcome of clinical care[5] or a better understanding of diseases.[6]

Increasingly, the value of genetic and health data for reuse beyond research is recognized.[7] Data sharing among scientific research organizations is of vital importance for advancing knowledge through supporting data analyses of greater statistical power, while also enabling validation and reproducibility of previously obtained results. In addition, secondary uses of data collection for policy development in public health and for regulatory purposes are emerging as promising approaches to utilizing existing health and genetic data to benefit society. Among other benefits, greater systematic secondary use would prevent new recruitment or further involvement of data subjects (patients or research participants), including those who may be 'fatigued' from ongoing calls for their data, and likely would help facilitate greater efficiencies in data flows and data management.

Enabling secondary uses for scientific research, which is a focus area of this article, has been actively pursued by the broader biomedical research community, not only through efforts to develop standards and best practices for making existing data FAIR (Findable, Accessible, Interoperable, and Reusable), but also by establishing data-sharing initiatives, platforms, and infrastructures.[8] Increasingly, these efforts include cross-border initiatives supported by governments of European countries, such as the Cancer Image Europe (EUCAIM) and the European 1+ Million Genomes (1+MG) Initiative.[9] These developments are being pursued concurrently with even larger-scale efforts by EU legislators to improve data sharing across the Union.

Indeed, in May 2022, the European Commission proposed the European Health Data Space (EHDS) Regulation, with a similar aim to EUCAIM and the 1+MG Initiative: to enable secondary use of existing valuable health data for a range of purposes, including research, healthcare reuse, and policy development. (As of the time of writing, the final text of the EHDS Regulation is not yet adopted.[10]) An explicit aim of the EHDS is to unlock data that are currently not available for secondary use as described in recitals 2 and 38 of the compromise text that was adopted based on a provisional agreement between the European Parliament and the Council.[11] While set up to solve data protection challenges, the proposal has not been without controversy, with some stakeholders raising concerns about the privacy-related implications of the proposed data uses.[12] All of these initiatives also align with the

2 Shahid M Shah and Rizwan Ahmed Khan, 'Secondary Use of Electronic Health Record: Opportunities and Challenges' (2020) 8 IEEE Access 136947.

3 Regina Becker and others, 'Secondary Use of Personal Health Data: When Is It "Further Processing" Under the GDPR, and What Are the Implications for Data Controllers?' (2023) 30 European Journal of Health Law 129.

4 Bernice S Elger and others, 'Strategies for Health Data Exchange for Secondary, Cross-Institutional Clinical Research' (2010) 99 Computer Methods and Programs in Biomedicine 230.

5 See eg, Stefan Larsson and others, 'Use of 13 Disease Registries in 5 Countries Demonstrates the Potential to Use Outcome Data to Improve Health Care's Value' (2012) 31 Health Affairs 1.

6 See eg, Shimaa A Heikal and others, 'The Impact of Disease Registries on Advancing Knowledge and Understanding of Dementia Globally' (2022) 14 Frontiers in Aging Neuroscience 774005.

7 Lisa M Federer and others, 'Biomedical Data Sharing and Reuse: Attitudes and Practices of Clinical and Scientific Research Staff' (2015) 10 PLoS One e0129506.

8 Maria A Rujano and others, 'Sharing Sensitive Data in Life Sciences: An Overview of Centralized and Federated Approaches' (2024) 25 Briefings in Bioinformatics bbae262.

9 Gary Saunders and others, 'Leveraging European Infrastructures to Access 1 Million Human Genomes by 2022' (2019) 20 Nature Reviews Genetics 693.

10 See European Parliament, 'Legislative Observatory: 2022/0140(COD): European Health Data Space' <https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2022/0140(COD)&l=en> (accessed 4 September 2024).

11 EU Council, 'Proposal for a Regulation on the European Health Data Space—Analysis of the Final Compromise Text with a View to Agreement' (18 March 2024) <https://www.consilium.europa.eu/media/70909/st07553-en24.pdf> (accessed 4 September 2024).

12 See eg, European Digital Rights, 'EU's Proposed Health Data Regulation Ignores Patients' Privacy Rights' and Position Paper' (6 March 2023) <https://edri.org/our-work/eu-proposed-health-data-regulation-ignores-patients-privacy-rights/> (accessed 4 September 2024); European Digital Rights, 'Joint Public Letter to EU Lawmakers on Patients' Rights in the European Health Data Space' (2023, April 13) <https://edri.org/wp-content/uploads/2023/04/Joint-public-letter-on-consent-in-EHDS-2.pdf> See also Ciara Staunton and others, 'Ethical and Social Reflections on the Proposed European Health Data Space' (2024) 32 European Journal of Human Genetics 498; Mahsa Shabani, 'Will the European Health Data Space Change Data Sharing Rules?' (2022) 375 Science 1357; Robin van Kessel and others, 'The European Health Data Space Fails to Bridge Digital Divides' (2022) 378 British Medical Journal e071913.

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

European Union (EU)'s 'European strategy for data',[13] the overarching objective of which is to enable the free flow of data across the Union, tapping into the economic (and social) value of the data for the benefit of all.

As is well known, the secondary use of sensitive data, which health and genetic data certainly are, must respect applicable privacy and data protection rules, laws, and regulations, and more rigorous rules apply to the processing of health and genetic data than 'regular' kinds of personal data, given their sensitive nature. In the EU, the General Data Protection Regulation (GDPR) is the principal component of the legal framework regulating the processing of personal data.[14] Various provisions of the GDPR are further implemented in the national data protection laws of the EU Member States. Together, the GDPR and its national implementations constitute the overarching privacy and data protection legal framework in Europe. This legal framework is complemented by sector-specific laws and regulations which, in the case of health and genetic data, include, among others: legal acts regulating the delivery of healthcare services, medical secrecy laws, and laws governing the conduct of biomedical research. It is also complemented by the relevant case law of the Court of Justice of the European Union (CJEU).

As we will argue in this article, in the EU, effective, responsible secondary use of health and genetic data is currently hindered by an interplay of complex, sometimes inconsistent requirements arising from the different legal frameworks or interpretations thereof. This makes the concept of an encompassing legal framework for the secondary use of these data, as endeavoured through the proposed EHDS Regulation, very attractive. However, as we go on to argue, the EHDS Regulation does not provide a 'silver bullet' to these fundamental data protection law issues. In general, the complexity and the magnitude of legal challenges tend to increase with the expanding scope of purposes for which these data are to be reused. This will be highlighted by comparing legal challenges, uncertainties, and non-compliance risks across the key contexts in which reuse holds significant promise.

The various hurdles for effective secondary use of health data have been outlined, among others, by a study commissioned by the European Commission[15] and in the case studies on barriers to cross-border sharing of health data for secondary use of the Joint Action TEHDAS,[16] a project funded through the EU4Health programme.[17] However, these reports do not provide an in-depth analysis of the legal situation under the GDPR for secondary use of health and genetic data. This article aims to fill this gap.

For brevity's sake, we will primarily focus this comparative analysis on the contexts of scientific research, and to a lesser degree, healthcare reuse of health and genetic data, while also touching upon other areas relevant to health-related data reuse. As regards the types of legal challenges in the way of cross-border secondary use of health data, we limit our analysis to the topic of legal basis within the meaning of Articles 6(1) and 9(2) GDPR. Since having a valid legal basis to process personal data is the *sine qua non* of GDPR compliance, we set out to examine which GDPR legal bases are available to the key actors involved in the health and genetic data reuse lifecycle. Following an in-depth analysis of the relevant GDPR legal bases, we make the case that the requirements regarding the legal basis for processing these data for secondary use vary significantly across EU Member States, which is problematic for pan-European and international data sharing initiatives. As a result of the dependency on legislative acts, as well as the varying nature of stakeholders, we find that there can be significant hurdles for secondary use and some actors can even be excluded from participation entirely. We then discuss how these challenges can be overcome through harmonized EU legislation and turn our attention to the legislative progress made thus far, primarily through the proposed EHDS Regulation. We note that the proposed EHDS Regulation is an important step in the direction of ensuring the availability of a valid legal basis under the GDPR. However, we also highlight that the GDPR legal bases created in the proposed EHDS Regulation are insufficient for many challenges of secondary use of data, as they leave out important steps within the data

13 European Commission, 'The European Data Strategy' <https://digital-strategy.ec.europa.eu/en/policies/strategy-data> (accessed 4 September 2024). Recent adoption of new European laws expanding the legal framework for the reuse of data, such as the Data Act and the Data Governance Act, are examples of the legislative developments aligned with the 'European strategy for data'. Sector-specific laws include the proposed EHDS Regulation, which will be further explored below.

14 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), OJ 2016 L 119/1. The GDPR defines 'personal data' as 'any information relating to an identified or identifiable natural person ('data subject')' (art 4 GDPR).

15 Johan Hansen and others, *Assessment of the EU Member States' Rules on Health Data in the Light of GDPR* (Publications Office of the European Union 2021).

16 Towards European Health Data Space (TEHDAS), 'Updated: TEHDAS Identifies Barriers to Data Sharing' (8 July 2021) <https://tehdas.eu/tehdas1/results/tehdas-identifies-barriers-to-data-sharing/> (accessed 4 September 2024).

17 European Commission, 'EU4Health' <https://commission.europa.eu/funding-tenders/find-funding/eu-funding-programmes/eu4health_en> (accessed 4 September 2024).

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

reuse lifecycle and are insufficient for data harmonization, which was also identified as a major hurdle. We conclude by outlining potential additional, solution-orientated European legislative efforts that could complement the EHDS Regulation by providing the possibility to create a valid GDPR legal basis for accessing harmonized personal data for secondary use.

To bring the situation for secondary use of health and genetic data and the potential future role of the EHDS therein into context, we begin our analysis by looking at how there is often a lack of a suitable legal basis for making health and genetic data available for secondary use and/or the user itself. This requires us to examine the prerequisites for secondary use under the GDPR before turning to analyse the strengths and weaknesses of the various legal bases available under Article 6(1) GDPR for secondary use of personal health and genetic data.

Prerequisites for secondary use under the GDPR

The overall data lifecycle, from data collection by an organization for its own purposes, to the data reuse by another party for the latter's purpose, can be summarized as follows in Figure 1.

We will be using the following three terms to denote the key parties involved in this processing chain: Data Provider, Data-Sharing Intermediary (DSI), and Data User. We note that these are not legal terms defined in EU law, but rather label the generic roles that we find useful for analysing the data reuse lifecycle outlined in Figure 1 because every legislation related to secondary use introduces terms for these roles in a specific context, which can mostly not be generalized.

We are aware that the EU legislative texts, such as the Data Governance Act[18] and the proposed EHDS Regulation, introduce similar—albeit not identical—roles involved in the secondary use of data. To avoid potential confusion, particularly in the latter parts of this article where we discuss the proposed EHDS Regulation and the Data Governance Act, Table A1 in the Annex offers a detailed overview of the relevant terms.

For the purposes of this article, we define the above three terms as follows:

Data Provider: an entity that originally collected health and/or genetic data for its own purposes and that subsequently wants to make the data systematically available for secondary use to Data Users (defined below) for specified purposes.

DSI: a party enabling the Data Provider to make the data systematically available to Data Users.[19] The DSIs can differ in their role, acting either as 'controller' or as 'processor' (key GDPR terms) for the subsequent disclosure of data to the 'Data User', depending on whether they have a decisive role with respect to data disclosure; they may or may not physically host the data, and they can also require certain models and formats of data and metadata to be observed. There can be obligations for Data Providers to submit data to a defined DSI or Data Providers can choose a DSI voluntarily as support. It is important to note that from an operational point of view, a DSI is *not* obligatory for the step of making data systematically available by Data Providers to Data Users. Data Providers can just as well decide to create their own repository to hold data for secondary use. Irrespective of the type and architectural patterns of DSIs, they all serve the same overarching purpose of helping make existing datasets broadly available to eligible external parties for secondary uses. Three exemplary types of such DSIs are described in Box 1: the European Genome Phenome Archive (EGA), the Danish National Genome Center, and the 1+MG Initiative. The qualification of a DSI in the EHDS and DGA are described in Table A1 of the Annex.

Data User: a party aiming to use, for its own purpose, the personal health data made available by the Data Provider. Each step along the data reuse lifecycle is associated with a set of processing operations on personal data. The exact nature of the processing operations will depend not only on the type and architectural pattern of the DSI that the Data Provider is utilizing (if any), but also various data governance aspects and other contextual factors, including the nature of the requests

18 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), OJ 2022 L 152 <https://eur-lex.europa.eu/eli/reg/2022/868/oj> (accessed 4 September 2024).

19 Of note, the use of a DSI is not strictly necessary in order for the Data Provider to make its data systematically available for third parties' research purposes. For example, the Data Provider may also set up an access governance framework internally, with a dedicated Data Access Committee (DAC) and online electronic means through which third parties may request access to data. The Data Provider may even implement an own secure data processing environment, enabling data analysis. Some large hospitals and biomedical research organizations may have the capabilities to do this. However, in practice, most Data Providers will lack the necessary resources to make their data collections systematically available in a meaningful way, ie enabling data discovery by the broader research community, transparently communicating access and use conditions, and streamlining the access governance process. See eg, Angela G Villanueva and others, 'Characterizing the Biomedical Data-Sharing Landscape' (2019) 47 Journal of Law, Medicine & Ethics 21. For these reasons, this article focuses on modalities involving secondary uses of data through DSIs.

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

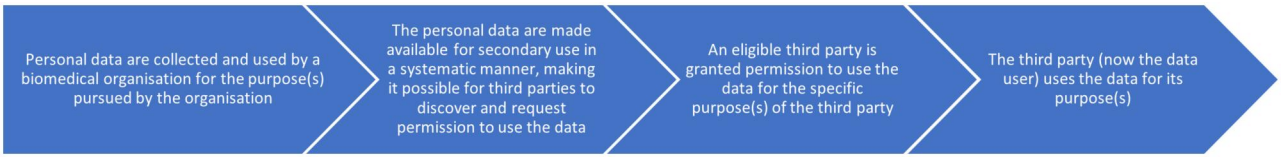

Figure 1. Data reuse lifecycle.

Box 1. Three examples of European DSIs with diverse governance or architectural design in the area of genetic/genomic data.

The European Genome-Phenome Archive (EGA)

A major European resource for the reuse of health and genetic/genomic data is the European Genome-Phenome Archive (EGA, https://ega-archive.org/). It has been described on the website as a 'service for permanent archiving and sharing of personally identifiable genetic, phenotypic, and clinical data generated for the purposes of biomedical research projects or in the context of research-focused healthcare systems'. The datasets included in the EGA are primarily intended for reuse in the research context. Research organizations making data collections available via the EGA retain full control over decisions regarding subsequent access to and use of the data collection they have contributed.

The 1+ Million Genomes Initiative

The 1+ Million Genomes (1+MG) Initiative (https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes) is a proposed DSI with the aim of making large-scale human whole genome sequence data together with phenotypic data widely available and reusable across Europe in a harmonized fashion. The Initiative is supported by 25 EU Member States, the UK, and Norway. It aims to establish a streamlined data-sharing infrastructure, alongside a harmonized governance process for facilitating genomic data-sharing. It is envisaged that the 1+MG Initiative will accept genomic and associated health-related data collections generated in various healthcare as well as research contexts. Subsequently, the 1+MG Initiative will make these collections available for a wide range of purposes, including, among other uses, research, healthcare provision, and health policy development.

Danish National Genome Database managed by the Danish National Genome Center

Established in May 2019 under a national Health Act (Bekendtgørelse af sundhedsloven, https://www.retsinformation.dk/eli/lta/2023/248), the Danish National Genome Center (NGC) 'supports the development of personalised medicine in collaboration with Danish research and healthcare organizations, and via international projects with stakeholders' (https://eng.ngc.dk/research-and-international-collaboration). It is tasked with managing the Danish National Genome Database, which includes granting access to the genomic data held in the Database, in a dedicated secure processing environment called NGC Cloud. Since 1 July 2019, Danish healthcare organizations generating clinical-grade genomic data are legally required to report or transfer the data to the NGC. These organizations play no role in downstream access and use decisions in relation to the datasets they contribute, with the Danish NGC assuming the decision-making role. Because of its focus on the development of personalized medicine, the NGC has specifically tailored its data access and secondary use processes to the context of research. However, considering that much of the genomic data held in the Danish National Genome Database is of clinical grade, it is foreseeable that streamlined processes and workflows for enabling other types of secondary uses will also be developed in the future.

from Data Users. For example, the precise set of processing operations may or may not involve actual transfer of data to a centralized repository, data cleaning, transforming the data to conform to a particular data standard mandated by the DSI, or subjecting the data to certain access and use conditions or restrictions. To aid with the discoverability of the data by prospective Data Users, it may also be required to generate descriptive statistics and other metadata pertaining to the underlying dataset. Similarly, at each subsequent phase in the data reuse lifecycle, various specific processing operations will be performed on the personal data, whose

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

precise nature will be influenced by various relevant contextual factors, including the architectural patterns of DSIs and the applicable data governance model.

Irrespective of how a data reuse lifecycle is structured, every processing operation carried out on the personal data must comply with the GDPR. In this respect, three overarching GDPR-related requirements are of particular importance. Namely, for each processing operation on personal data, the following three elements must be defined:

(i) the purpose(s) for which the processing is performed;
(ii) the data controller(s) for the processing; and
(iii) a valid GDPR legal basis in which the processing is grounded.

Various commentators have covered in depth the challenges secondary use of health and genetic data raises in respect of the first two components, namely, the purpose of the processing and data controllers involved for the processing activities.[20] In this article, we are interested in discussing the third component, which remains underexplored in the literature. By legal basis, we mean meeting at least one of the six conditions listed under Article 6(1) GDPR, in conjunction with one of the ten exemptions under Article 9(2) GDPR where special categories of personal data may be lawfully processed (this includes health and genetic data).[21] It is the responsibility of the controller to demonstrate that every processing operation has a valid legal basis and in the case of special category personal data, a valid exemption. Importantly, the validity of a GDPR legal basis must be assessed in view of all of the above elements: the nature and context of processing; the specific purpose(s) for which the processing is being carried out; and the identity, legal structure, and the jurisdiction of establishment of the controller.

Legal basis is a critically important concept for compliance reasons. A missing or insufficient legal basis is the single most frequently cited violation in GDPR fines issued by supervisory authorities.[22] Yet, as the next section will illustrate, in a data reuse lifecycle where data are shared for secondary use (Figure 1), ensuring that *all* the controllers have a valid legal basis is in practice a significant challenge and raises profound questions about the adequacy of the existing EU data protection regime. This is where we think the 1+ Million Genomes Initiative and other pan-European data-sharing initiatives run into legal trouble, and where to date there are no effective legislative solutions, as we now proceed to demonstrate. Even the EHDS, as we proceed to argue below, will not provide a legal basis for all processing elements related to secondary use.

Challenges with the existing GDPR legal bases for processing health and genetic data for secondary use

In this section, we illustrate the challenges in identifying a valid legal basis for secondary use of health and genetic data. We focus our analysis primarily on the context of scientific research, but it can be seen as having some application to other contexts, such as health care, and indeed, we consider this in the analysis below. In our assessment, the relevant legal bases to consider are: (i) consent, (ii) legal obligation, (iii) performance of a task in the public interest, and (iv) legitimate interest. We explore each of these in turn.

The challenge with consent: GDPR Articles 6 (1)(a) + 9(2)(a)

Many in the health research community might think that consent would be an appropriate legal basis to facilitate the secondary use of health and genetic data. This is due to the conflation of consent as a legal basis to process personal data with the ethical norm to secure informed consent for participation in health research. This widespread confusion in the research community has also been dubbed the 'consent misconception'.[23] As obtaining a research participant's informed consent is a

20 See eg, Regina Becker and others, 'Purpose Definition as a Crucial Step for Determining the Legal Basis under the GDPR: Implications for Scientific Research' (2024) 11 Journal of Law and the Biosciences lsae001; European Data Protection Board, 'Guidelines 07/2020 on the Concepts of Controller and Processor in the GDPR' (2020) 07/2020 <https://edpb.europa.eu/our-work-tools/documents/public-consultations/2020/guidelines-072020-concepts-controller-and_en> (accessed 4 September 2024); Inger Johanne Bakken and others, 'The Norwegian Patient Registry and the Norwegian Registry for Primary Health Care: Research Potential of Two Nationwide Health-Care Registries' (2020) 48 Scandinavian Journal of Public Health 49.

21 As stated in Article 9(1) GDPR, special categories of personal data include: 'personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership […] genetic data, biometric data […] uniquely identifying a natural person, data concerning health […] data concerning a natural person's sex life or sexual orientation'.

22 Marlene Saemann and others, 'Investigating GDPR Fines in the Light of Data Flows' (2022) 2022 Proceedings on Privacy Enhancing Technologies 314.

23 Edward S Dove and Jiahong Chen, 'Should Consent for Data Processing Be Privileged in Health Research? A Comparative Legal Analysis' (2020) 10 International Data Privacy Law 117; Eugenijus Gefenas and others, 'Controversies between Regulations of Research Ethics and Protection of Personal Data: Informed Consent at a Cross-Road' (2022) 25 Medicine, Health Care and Philosophy 23.

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

common ethical (and, in some cases, legal) requirement, many would be inclined to readily assume that the processing of personal data is also based on consent within the meaning of Article 6(1) GDPR. However, as we proceed to argue, this legal basis in fact raises a number of conceptual and practical problems that make it a poor option. We start by looking at the principal requirements for consent as a means to then illustrate its problems when applied to secondary use in the research context, particularly *via* the vehicle of 'broad consent', which is ambiguously referenced in recital 33.

Principal requirements for consent

Consent as a legal basis in the GDPR has been designed to give data subjects the greatest extent of control over their data, ostensibly effectively empowering them to decide how their data may or may not be processed.[24] Consent must be 'freely given, specific, informed and unambiguous' (Article 4(11)) and it must be possible for the data subject to withdraw the consent at any time (Article 7(3)). The legislator has taken considerable care to ensure that consent is obtained for a well-defined purpose. The legislator has done so by requiring that consent be given for:

- A *specific* purpose, as per Article 6(1)(a). This implies that a valid consent can only be given for an individual purpose and not for bundled purposes, an interpretation also supported by recitals 32 and 42 GDPR (eg, recital 32: '[when] the processing has multiple purposes, consent should be given for all of them').
- A *specified* purpose, as per Article 9(2)(a). This means that each distinct purpose for which the controller intends to process special category personal data must be explicitly delineated, allowing the data subject to decide whether to consent to processing for that purpose.

Additionally, Article 7(1) states that '[w]here processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing his or her personal data'. In the context of complex health or genetic data processing chains involving multiple controllers, this means that each controller relying on consent as the legal basis must be able to demonstrate that explicit consent has been obtained from the data subject that is valid with respect to the processing operations pursued by this controller. This should also be read in light of recital 42 that for consent to be informed, in keeping with the definition of Article 4(11) GDPR, the controller needs to be known to the data subject.

Recital 33 as a potential gateway to broad consent?

In the context of scientific research, potential concessions on the specificity of consent can be considered. In particular, recital 33 GDPR states the following (we quote in full):

> It is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection. Therefore, data subjects should be allowed to give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose.

Recital 33 has been interpreted to allow for broad consent (a concept imported from research ethics) to future secondary use. This interpretation has found considerable appeal among the research community, owing to its research-enabling implications. However, a more detailed analysis demonstrates that there are considerable limits to this interpretation. Namely, the broad consent that appears to be enabled by recital 33 GDPR may be limited in scope to the *primary purposes* pursued by a particular controller when collecting the data, thus excluding any future secondary use.

This interpretation is grounded in the following considerations of the recital's text.

'to fully identify'. The reference in recital 33 GDPR is to a purpose not *fully identified* at the time of data collection by a controller. This suggests that the purpose needs to be identified to a certain extent, and certainly one far more than *de minimis*. It stands to reason that the purpose should be sufficiently detailed to inform the design of the initial processing by the controller. In other words, the not-yet-fully-identified purpose must nevertheless be formulated in a manner that, at a minimum, allows for designing the data collection step. Designing data collection, in turn, entails defining the categories of data subjects and to a certain extent the data types to be obtained from the data subjects. Additionally, the design of the data collection should entail clear indications as to how the processing continues beyond the collection, which can then lead to a

24 Javed Ahmed and others, 'GDPR Compliant Consent Driven Data Protection in Online Social Networks: A Blockchain-Based Approach' in *2020 3rd International Conference on Information and Computer Technologies (ICICT)* (IEEE 2020) <https://ieeexplore.ieee.org/document/9092226/> (accessed 4 September 2024); Dove and Chen ibid.

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

specification of the purpose as required. Controllers always have to comply with data protection by design and default requirements of Article 25 to the extent a purpose is defined. In other words, the 'not yet fully defined purpose' is nevertheless what informs the collection and subsequent processing (eg, setting up the data resource, building up a cohort) and needs to be explained to data subjects.

Future secondary use, on the other hand, is yet unspecified and there are typically no processing operations associated with it at the time of data collection. Recitals 32 and 42, together with Article 7(1), suggest that for a valid consent, the delineation of relevant processing operations is a crucial element, further suggesting that a purpose that is not directly linked to specified processing operations cannot be valid. A purpose must, whenever possible, always be identified (and communicated to the data subject) before a processing activity commences.

In their interpretation of recital 33, the European Data Protection Board (EDPB) also refer to a primary research project in their 'Guidelines 05/2020 on Consent' and do not consider undefined future projects: 'For the cases *where purposes for data processing within a scientific research project* cannot be specified at the outset, Recital 33 allows as an exception that the purpose may be described at a more general level' (emphasis added).[25]

'Not possible'. The GDPR typically foresees concessions where the objectives of the processing could otherwise not be reached under the applicable circumstances. There is a fundamental difference between primary and secondary use of data with respect to processing by the controller. Where it is impossible to fully identify the purpose of the primary processing at the time of collection, such impossibility must stem from the nature of the processing. An example of an intrinsic impossibility to fully identify the purpose is a population cohort where the downstream precise research questions depend on the health and disease progression of the research participants. There is an intrinsic impossibility to fully identify the purpose at the time of collection. For secondary use in future research projects on the other hand, the impossibility to identify the purpose is due to the fact that the projects are not yet defined. This is a temporal rather than a fundamental problem that does not affect the lawfulness of the collection. For downstream processing at a later stage, the legal basis has only to be established at the time when this processing commences.

The requirement for consent to be given for specific purposes means a consent would not be valid unless the purpose is specified at the time of collection. This would also apply to situations where it is inherently impossible to fully identify the specific purpose, thus rendering the consent invalid. Considering that the GDPR makes concessions for situations that are otherwise not possible, this could explain the reason to open up consent to research that cannot be specified entirely at the outset. In the case of secondary use, however, the reason to compromise on the specificity of consent at the time of collection would rather be a convenience argument, avoiding the subsequent re-consenting for new research projects. It does not seem to be likely that 'effort' can be a compelling argument to weaken the core principle of specificity of consent. This consideration suggests that the 'not possible' refers to an intrinsic impossibility, rather than temporal inability, to specify in advance the purposes for future research projects.

Further processing. Future research-related activities that are not defined at the time of collection always give rise to further processing, that is, processing for a (specific) purpose that was not driving the collection. Further processing takes place for a new purpose by the same controller.[26] For example, in the case of the data reuse lifecycle depicted in Figure 1, processing operations falling under the step 'making data systematically available' would constitute further processing by the Data Provider.

The challenge that not all future purposes can be identified at the time of collection is not specific to research; it may also apply to other secondary uses. Once again, this supports the interpretation that the exceptions permitted by recital 33 should be understood narrowly, that is in relation to the primary use of data in the research context where the inherent nature of the research is such that it prevents the controller from fully identifying the purposes in a GDPR-compliant manner, at the time of data collection.

In the case of primary purposes, consent may be given at the time of collection to a not yet fully identified purpose for areas of research that split into

25 The European Data Protection Board, 'Guidelines 05/2020 on Consent under Regulation 2016/679' (2020) 05/2020 (para 156) <https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf> (accessed 4 September 2024). See also 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' (2021) 07/2020 (para 26) <https://edpb.europa.eu/our-work-tools/our-documents/other-guidance/edpb-document-response-request-european-commission_en> (accessed 4 September 2024).

26 Becker and others (n 3).

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

different, more precise research questions, all covered by that consent, as there is a direct line between the collection and the downstream projects for which the collection was designed. Yet, for further processing undertaken at a later stage, ie, where processing for that purpose only starts at a later stage, no necessity can be argued that in advance, broad consent is needed. Consent could be obtained at the start of the processing for further processing, which means re-consenting would be required, in line with the GDPR principle that consent has to be given for all purposes (eg, recital 32 GDPR).

Application of the above considerations to data sharing for secondary use

In light of our analysis above, we find that the secondary use of existing personal data for research by the scientific community is a downstream processing operation that does not lead to the design of the collection. As shown in Figure 1, several successive purposes need to be considered in the corresponding secondary use data processing lifecycle.

The step of making data legally available to be offered for secondary use in research can be a conscious decision by the initial controller that collects the data from the data subjects. The 'how', ie, the essential means of processing determined by the controller, refers to the substantive technical, operational, and organizational elements of the data access governance framework. The 'what' of the processing concerns the categories of purposes for permissible data reuses, as known at the time. Hence, prior to commencing processing operations towards making the data widely available, the controller must obtain a specific and informed consent from the data subject, encompassing these informational content elements. More specifically, this consent will ensure that data are made FAIR in line with the information provided at the time of collection. Such consent is specific for associated processing operations, which can even include a transfer to a DSI that subsequently takes over the responsibility to take the access decisions as long as such transfer is sufficiently described in the consent. However, consent is not only required to be specific; it must also be freely given. This can limit the possibility where data are obtained in a healthcare context and there is a perceived or real imbalance of power between the data subject and the controller.[27] This is often the case with, say, paediatric rare disease patients, where parents are desperate to find a diagnosis and/or cure for their child. There are also other possibilities of an imbalance where vulnerable subjects such as asylum seekers are implicated.[28]

Consent covering the disclosure and processing for downstream research projects or to healthcare professionals for reuse in health care becomes less straightforward. In research, every time access is provided, this constitutes a distinct purpose of processing personal data pursued by the controller deciding on whether to grant access, where each scope of data disclosed must be specified by the respective research project.[29] Therefore, in the strict interpretation of consent as introduced in Articles 6(1)(a) and 7, and supported by recitals 32 and 42, a separate consent is required for each purpose, in this case for each data disclosure for a Data User's research project. The interpretation that each subsequent disclosure or permission to reuse the data for a particular research project requires a specific consent by the data subject has led to the development of digitally enabled dynamic consent solutions.[30] Dynamic consent, for example, means that data subjects must consent to the reuse of their personal data in each research project pursued by a third party.[31] This approach, alongside the digital dynamic consent management tools necessary for supporting a demonstrably GDPR-compliant consent, has been gaining traction in some European countries, such as Italy and Malta.[32]

It might be argued that in those cases where a change of controller takes place between making the data available by the initial controller and the DSI being responsible for the data sharing, the DSI has collected the data for the purpose to contribute the data to Data Users' research projects. This could be seen as a primary purpose of the DSI that is not yet fully identified at the time of collection, as the individual project will only be defined later. Strictly speaking, as the DSI is not participating in the actual research, it can be questioned if recital 33 is

27 European Data Protection Board, 'Opinion 3/2019 Concerning the Questions and Answers on the Interplay between the Clinical Trials Regulation (CTR) and the General Data Protection Regulation (GDPR) | European Data Protection Board' (23 January 2019) <https://edpb.europa.eu/our-work-tools/our-documents/opinion-art-70/opinion-32019-concerning-questions-and-answers_en> (accessed 4 September 2024).

28 Christopher Kuner and Massimo Marelli (eds), *Handbook on Data Protection in Humanitarian Action* (ICRC 2017).

29 Becker and others (n 3).

30 DNV GL, Group Research and Development, Precision Medicine Program, 'Dynamic Consent in Clinical Genetics: Implementation Barriers' (2021) <https://www.dnv.com/research/healthcare-programme/dynamic-consent-whitepaper.html> (accessed 4 September 2024).

31 Harriet JA Teare, Megan Prictor and Jane Kaye, 'Reflections on Dynamic Consent in Biomedical Research: The Story So Far' (2021) 29 European Journal of Human Genetics 649.

32 Deborah Mascalzoni and others, 'Ten Years of Dynamic Consent in the CHRIS Study: Informed Consent as a Dynamic Process' (2022) 30 European Journal of Human Genetics 1391; Nicholas Mamo and others, 'Dwarna: A Blockchain Solution for Dynamic Consent in Biobanking' (2020) 28 European Journal of Human Genetics 609.

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

applicable at all to the DSI as the DSI itself does not define the research and does not participate in the research.[33] But even if the support of research is assumed to be included in the intention of the recital, the above considerations of an inherent inability to fully identify the purpose at the time of collection apply to the DSI. Therefore, a DSI might obtain consent from the data subjects for contribution of data to Data Users' projects in defined areas of research within the meaning of recital 33. However, even if recital 33 applied to the processing of the DSI when collecting the data, that is when holding them in their dedicated database, the recital is silent about what happens where a purpose becomes fully identified.

The EDPB suggest in their Guidelines 05/2020 at paragraph 159 that re-consenting should be pursued before 'the next stage begins'. At paragraph 161, transparency on repeated information also is proposed as a safeguard. Considering the rather binary step of knowing the specific data use for which data will be disclosed, paragraph 159 seems to be more applicable. This is to be seen in view of the definition of consent as a specific, informed, and unambiguous indication with a statement or clear affirmative action that signifies the agreement to the processing of data related to the data subject. This requirement does not appear to be compatible with a situation where a controller relies on information provision only with respect to the data subject and provides only an opt-out as signal from the data subject that they do not agree. This is because it will be difficult for the controller to demonstrate that the data subject has received the information and is in actual agreement (indeed, the CJEU case C-61/19 *Orange România SA* holds that that without clear feedback, the consent does not fulfil the 'unambiguous' condition). In particular, once data are disclosed, the consent cannot be withdrawn retroactively. This almost binary step to process data for a finally identified purpose seems to tip the balance towards a necessity of an affirmative action by the data subject.

For healthcare reuse, it could be argued that 'informing the diagnosis of a similar patient' or 'informing the therapy of a similar patient' is a specific enough purpose. However, a closer look makes it clear that there can still be differences depending on the kind of question that needs to be answered, such as differences in the processing operations and the data to be consulted, among others.

Data Users who gain access to the data for their research projects or for healthcare reuse must again obtain their own consent: for consent to be valid, the identity of the controller must be known (recital 42). Importantly, the Data User cannot obtain such a consent without the support of the initial controller or the DSI, whichever acts as the controller for granting the permission to use the data. This is because the Data User has no means for re-contacting the data subject, rendering the Data User's ability to obtain a GDPR-compliant consent contingent upon whether the upstream controller (either the DSI or the Data Provider, as applicable) can support the Data User in this regard. Where the upstream controller has the means, and agrees to enable the reconsenting process, the upstream controller must ensure that the new consent is obtained for the specific purpose pursued by the Data User, and that the data subject is informed about the relevant controller's (ie, the Data User's) identity as part of the consenting process.

In sum, the limited possibility of a broad consent under the GDPR makes the reliance on the legal basis under Articles 6(1)(a) and 9(2)(a) conceptually difficult and an impractical solution that will easily lead to consent fatigue and thus a high attrition rate and a bias in the available data.[34] However, even if broad consent were possible, we note that relying on consent as a legal basis limits continuous availability of data for future secondary uses. This is because, for a consent to be valid, the data subject should have the right to withdraw the consent at any time.

Collectively, the aforementioned challenges of consent for processing health and genetic data make it an impractical legal basis in the context of secondary data use lifecycle. This conclusion, as we noted previously, runs counter to the widely held view among the scientific research community that recital 33 permits a broad consent at the time of data collection, which remains valid for future further processing, that is, without the need to obtain a new consent.

This conclusion necessitates reliance on alternative legal bases to ensure lawful processing of personal data throughout parts of the data reuse lifecycle. However, although these alternative legal bases exist, we go on to argue below that their availability or suitability is currently limited across EU Member States.

The challenge with legal obligation: GDPR Article 6(1)(c)

Legal obligation is applicable as a legal basis when the processing is required by a Union or Member State law to which the controller is subject. There can be no discretion on the side of the controller whether or how to

33 We shall engage further with this argument in a future manuscript.

34 Teare, Prictor and Kaye (n 31).

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

comply with this law: the law must provide the purpose,[35] which must be sufficiently specific as to which processing is or is not included.[36]

In theory, relying on Article 6(1)(c) as a legal basis would be the most straightforward route given the clearly prescribed steps the initial controller would need to follow. Where a legal obligation is established, the corresponding law should provide automatically for the exemption under Article 9(2) and should therefore also include suitable and specific safeguards. Currently, such obligations exist in the context of population health registries, where healthcare organizations such as hospitals and clinics are required to transfer certain types of patient data to designated entities (health registries). A similar situation arises where countries establish national or regional genome centres to which genome-sequencing organizations are legally required to submit genomic and related health data of their patients (see eg, the Danish NGC in Box 1). Such registries often have a primary mission of supporting personalized medicine within the healthcare system but typically also receive a legal mandate to contribute the data to research projects of Data Users. It is to be noted that healthcare reuse of data in such registries is still uncommon.

The corresponding legal basis for the disclosure to Data Users in these cases is not, however, Article 6(1)(c). The reason is that data disclosures to Data Users follow a defined data governance that still requires a consideration and decision by the repository/health registry as to whether to disclose the data. Legal obligation is, therefore, not a suitable legal basis to cover the contribution of data to individual research projects. These considerations are also reflected in the legal bases assumed in the proposed EHDS Regulation, as described below.

The same consideration applies to the Data Users themselves: there is no legal obligation on the part of the Data User to perform a prescribed research study. Rather, the Data User has considerable discretion to determine the research question even if the general areas of research may be circumscribed as part of the mission given to the organization by law. Also, the processing of personal data for the provision of health care is a task assigned rather than an obligation, as it is not possible to legally predetermine the processing to the level of detail required for Article 6(1)(c) in combination with Article 6(3).

The challenge with performance of a task in the public interest: GDPR Article 6(1)(e)

Similar to Article 6(1)(c), to rely on a task in the public interest as a legal basis, the basis of the processing has to be laid down in either Union or Member State law to which the controller is subject. However, in this case a *task* rather than a purpose is prescribed by the law, which means there is discretion as to the purposes that are covered by the law.[37] The task is often part of the mission given to an entity by a Union or Member State law. This may cover private bodies operating in health care as well as private bodies supporting governmental policy development. In all cases, the controller must be able to demonstrate that the processing is necessary to fulfil the task given by law.

The possibility to demonstrate the necessity of the purpose for the task or mission assigned is crucial, in particular where the initial controller who collected the data for its primary purpose, such as a specific research project, intends to make the data subsequently available for Data Users. Many public research stakeholders (eg, government-funded research organizations), for example, only have a mission to pursue research. However, making data widely available through their own repositories or a DSI, or contributing data to another researcher's project without the Data Provider's involvement in the project, does not, in our view, constitute research, but rather *research support*.[38] Such research support, however, is often not acknowledged as the part of research mission given by law. This observation is supported by the key performance indicators applied to the organizations, which are extramural funding, publications, and collaborations.

A direct sharing of data with Data Users is included in the mission of some organizations as already illustrated above: these are dedicated registries in health care, formally established repositories such as genome centres, and biobanks. Some EU Member States have also adopted additional laws to enable the establishment of research-facilitating DSIs at the national level. Notable examples include dedicated research repositories tasked with administering health data to enable future data reuses, including for research (eg, the Danish

35 See Article 6(3) GDPR.

36 See also Article 29 Working Party, Opinion 06/2014 on the notion of legitimate interests of the data controller under art 7 of the Directive 95/46/EC: 'Further, the legal obligation itself must be sufficiently clear as to the processing of personal data it requires. … The controller should not have an undue degree of discretion on how to comply with the legal obligation.' Article 29 Data Protection Working Party, 'Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC' (2014) WP 217 <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf> (accessed 4 September 2024).

37 See also Article 6(3) GDPR: 'The purpose of the processing shall be determined in that legal basis or, as regards the processing referred to in point (e) of paragraph 1, shall be necessary for the performance of a task carried out in the public interest.'

38 s 4, Data Protection Act 1050/2018 2019 (Finland).

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

NGC), as well as national permit authorities with a legal mandate to disclose personal data held by certain organizations within the country to external requesters for the latter's research use.[39] These and other DSIs established under a national law operate based on Article 6 (1)(e) when deciding on whether to permit the processing of personal data by a particular prospective Data User. Similar to Article 6(1)(c), the relevant national law serves as an exemption for the DSI from the controller's obligations under Article 9(2) in relation to data disclosure for the Data Users' purposes. In these more recently established legislations, a wider scope of purposes is foreseen, possibly indicating a new legislative approach towards legitimizing data reuse.[40]

However, while these national authorities have a valid legal basis to contribute the relevant data directly to the Data User, they may not be able to lawfully make the same data available through another DSI, which can become relevant if data are to be disclosed in a harmonized fashion through a pan-European DSI. Depending on the national authority DSI's precise mission, as prescribed by the relevant law, engaging another DSI may not be allowed unless it acts as a sub-contractor for the national authority's tasks. However, even where this is in principle permissible, making data more broadly available through another DSI may not be possible for practical reasons: for example, the data governance framework of the recipient DSI may be deemed incompatible with the national authority DSI's own data governance policies and/or compliance demonstration needs. This has particularly negative consequences where pan-European DSIs are envisaged that provide access to cross-border datasets that build on harmonized collections and that follow a common data governance framework, as envisaged, for example, in the 1+MG Initiative. The national authority DSIs' lack of a suitable legal basis to participate in such cross-border initiatives has a detrimental impact on Member States' joint effort aimed at creating large-scale, harmonized highly valuable datasets that can be efficiently accessed through a single access mechanism. These pan-European initiatives are vitally important to help change the current fragmentary landscape of European health data resources, whereby researchers interested in using them must navigate a multitude of dissimilar access applications, each subject to a different data governance framework dictated by locally applicable rules, national laws, and additional requirements determined by the controller for data disclosure.[41]

With respect to Data Users within the public sector, the legal basis under Article 6(1)(e) is a seemingly straightforward path for the pursuit of their purposes as these purposes are typically explicitly reflected in the mission given to them by law. Nevertheless, it is worth studying the two considered use applications of research and healthcare reuse separately.

Where pursuit of research has been incorporated into the mission of an organization, researchers from the organization can rely on a task in the public interest legal basis. On the other hand, other types of entities (including commercial companies) will not be covered by the same legal mandate in most countries, effectively preventing them from relying on Article 6(1)(e) legal basis. A notable exception is Finland, where scientific research can be pursued under Article 6(1)(e) in general, independent of the nature of the controller.[42]

Nevertheless, the broad application of a research mission or of scientific research in the public interest does not extend to the processing of special categories of data, such as health and genetic data. It is therefore necessary that the special category processing exemption under Article 9(2) be established separately, in addition to the Article 6(1) legal basis. While most EU Member States have implemented the scientific research exemption under Article 9(2)(j) GDPR in national legislation to allow the processing of special categories of data for such purposes, many Member States still also require a (research ethics) consent to be obtained where possible, which can only be exempted based on an express approval by a competent ethics committee.

These and other barriers to the access and/or use of health and genetic data for research purposes often stem from the provision in the GDPR in Article 9(4) that Member States can further limit the processing of health and genetic data beyond requiring the safeguards established in legislative acts to allow the processing of special categories of data. Additionally, as noted earlier, some Member States may require obtaining consent for the data reuse, either in the form of informed consent in the sense of a research ethics norm, or—more

39 Such dedicated national permit authorities include, for example, Findata in Finland and the French Health Data Hub in France. See Balint Ferencz and Bettina Buki, 'Three Ways of Secure Data Reusability in Europe: German Research Data Centres, Finnish Findata and the French Secure Access Data Centre' (2022) 2022 ELTE Law Journal 81.

40 See eg, s 2 of the Finnish Act on Secondary Use of Health and Social Data foresees use for statistics, scientific research, development and innovation activities, education, knowledge management, steering and supervision of social and healthcare authorities, and planning and reporting duty of an authority <https://stm.fi/documents/1271139/1365571/The+Act+on+the+Secondary+Use+of+Health+and+Social+Data/a2bca08c-d067-3e54-45d1-18096de0ed76/The+Act+on+the+Secondary+Use+of+Health+and+Social+Data.pdf?t=1559641328000> (accessed 4 September 2024).

41 Thijs Devriendt and others, 'An Agenda-Setting Paper on Data Sharing Platforms: EuCanSHare Workshop' (2021) 1 Open Research Europe 80.

42 See n 38.

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

problematically and with fewer exemption possibilities—within the meaning of the GDPR legal basis. In some cases, both types of consent may be needed. For instance, a recent review of Member States' national legal frameworks has found that countries in Southern and Eastern Europe generally have a stronger emphasis on consent.[43] Therefore, even where a Data User can rely on the Article 6(1)(e) GDPR legal basis to pursue research, this Data User may nevertheless encounter significant barriers to accessing and/or using the relevant data where the upstream party acting as the controller for data disclosure operates under a more restrictive legal framework.

Data Users from the healthcare context, ie healthcare professionals, may act under Article 6(1)(e) when delivering health care to patients, but in some countries, also Article 6(1)(b) (ie, performance of a contract) is a valid legal basis.[44] Nevertheless, healthcare provision is anchored in the national or regional healthcare legislation and is therefore also covered by the exemption under Article 9 (2)(h), including the professional secrecy obligation required under Article 9(3). However, a closer inspection reveals that very often medical secrecy and the safeguards of the implementation of Article 9(2)(h) are limited to the direct relationship of a healthcare professional with a patient. The possibility of a healthcare professional processing data of patients where this processing is not related to their care, but rather another patient's care, is not explicitly covered in most Member State laws. This gives rise to legal ambiguities as to whether healthcare professionals can lawfully use the personal data of patients not under their care in order to facilitate the diagnosis or treatment of their own patients. While this legal ambiguity generally creates room for interpretation, such interpretive flexibility is reduced where Article 9(2)(h) is only invoked through specific acts on primary healthcare provision that do not cover unrelated personal health data.[45] Some Member States have realized the associated weaknesses of the system and introduced specifically that personal data of other patients can be shared and used in the healthcare process.[46]

The challenge with legitimate interest: GDPR Article 6(1)(f)

To rely on Article 6(1)(f) as a legal basis, the controller must be able to establish its own legitimate and present interest or a legitimate and present interest of a third party that it is serving. Similar to 'task' under Article 6 (1)(e), an interest is broader than a purpose, and many purposes can be pursued under an interest. However, the interest must be real and present, corresponding to current activities or a benefit expected in the immediate future; interests that are too speculative will not be sufficient. The controller further needs to demonstrate in a balancing test that its own interest does not unduly impact on the interests and fundamental rights of the data subject. A positive outcome of the balancing test is likely where data are reused for purposes with societal benefits such as research and healthcare reuse, and assuming that pseudonymized data are processed under appropriate technical measures and safeguards, with a high level of transparency (ie, adequate and timely information provision to the data subject with a possibility to object or to otherwise opt out of the processing). However, this legal basis is not available to public authorities in the execution of their tasks.[47]

A legitimate interest in making data available for secondary use can be established for research stakeholders who need to make data from their research sustainable and reusable by the research community (eg, based on organizational or funder policies). But this would only cover a research reuse and is therefore limited. The wider scope of purposes as envisaged in certain broad-scope, cross-border DSIs, such as the 1+MG Initiative, would not be covered as other purposes such as healthcare reuse or policy development are not in the interest of a research organization. For stakeholders in health care, their own interests will, in most cases, be difficult to establish.

An interesting option is to consider deriving an Article 6(1)(f) basis from the DSI's mission to provide data for secondary use, which constitutes a present and legitimate interest in receiving the data. In this case, it is not the initial controller's own interest that justifies the processing, but rather that of the DSI as a third-party recipient, which aims to provide data to Data Users in research and health care. A change from the initial controller's interest to the DSI's interest as the motivation for the transfer must lead to a change in the purpose specification. The purpose would be different in this case: as opposed to making the data available through a certain DSI by the initial data controller, the purpose pursued by the initial data controller would need to be

43 Olga Tzortzatou-Nanopoulou and others, 'Secondary Use of Data for Research across Europe: In Search of a Minimal Common Denominator' (forthcoming paper on file with authors).

44 Hansen and others (n 15).

45 See eg, Poland: (i) Act of 10 May 2018 on the Protection of Personal Data 2018; and (ii) Act of 27 October 2017 on Primary Healthcare 2017.

46 See eg, Norway, Act of 2 July 1999 Relating to Health Personnel: Lov om helsepersonell m.v. (helsepersonelloven) 1999.

47 Article 6(1) GDPR: 'Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the performance of their tasks.'

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

framed along the lines of supporting the DSI in providing data to Data Users for defined purposes. The interest in the availability of the data in this case would be with the DSI and no longer with the initial controller. While this legal basis should be available to many stakeholders, including private and public research stakeholders as well as various healthcare providers, it will not be available to those entities that qualify as public authorities under national law and that have collected the data in the performance of their tasks as stated in Article 6 GDPR. These entities will have to rely instead on Article 6(1)(c) or Article 6(1)(e).

Another limitation of this legal basis is the necessity to additionally have a recourse to a suitable Article 9(2) exemption from the general prohibition on the processing of health and genetic data. Whereas processing of personal data in accordance with Article 6(1)(f) does not require the controller to rely on another law outside the GDPR, this is not the case with respect to the corresponding Article 9(2) exemptions where special categories of personal data are concerned. Hence, in order to rely on a relevant Article 9(2) exemption, such as Article 9(2)(j), in conjunction with Article 6(1)(f), the controller must be able to process special categories of data based on Member State or Union law *on top of* the legal basis. Such a law must cover the possibility of making special categories of personal data available for secondary use in the relevant domains. Many Member States lift the prohibition in a broad way for research purposes, which could include making data available for future use in research.[48]

Yet, questions remain about just how broadly 'scientific research' ought to be interpreted (despite recital 159's declaration to adopt a broad interpretation): can it cover data sharing for a research purpose even if that processing activity in itself is not generating knowledge or contributing to generalizable knowledge? Some countries provide more clarity as they refer more specifically to the conduct of research which does not seem to include activities relating to supporting research pursued by other parties (eg, the Swedish Ethics Review Act concerning research involving humans).[49] In addition, for health and genetic data, many Member States still require consent and have obligatory ethics committee approvals or exemptions in place where consent is not possible. As soon as secondary use purposes other than research are envisaged, though, there is little to be found in the current legislation.[50]

In sum, the necessity to have parallel legislation and the deficiency of such legislative acts means that legitimate interest will not provide a viable option to make health and genetic data available for secondary use.

The same considerations apply to the initial controller's or, where applicable, the DSI's contribution of health and genetic data to the Data User's purposes: where the initial controller or the DSI has established a repository and therefore an interest in contributing health and genetic data to Data Users' secondary use purposes, Article 6(1)(f) is a possible option, provided that the controller is not a public authority. However, the limitations stemming from the uneven availability of relevant laws that would enable the controllers to make use of an Article 9(2) GDPR exemption renders this possibility inefficient.

The situation eases to the extent that the exemption from the prohibition of Article 9(1) is covered for the direct pursuit of research in most Member States. However, this does not always include private stakeholders, whereas public bodies may find it difficult to justify this path in view of the limitations on the use of Article 6(1)(f) by public authorities. Moreover, other barriers and/or prerequisites to the use of personal data, such as pre-approvals by relevant committees and/or other designated bodies, will still apply.

For health care, once again, the situation is even less straightforward, as the implementation of Article 9(2)(h) into the relevant laws is largely limited to the context of medical professionals' interactions with their own patients. Therefore, in many scenarios involving the reuse of health and genetic data in the healthcare context, the Article 9(2)(h) exemption will not be valid.

Where data are not shared from a defined repository with a data governance either established by the initial controller itself or a DSI that acts as the controller for sharing the data with the Data User, the legitimate interest changes to the interest of the user as a third party. Otherwise, however, the conclusions above apply in the same way.

This situation is expected to be changed with the creation of the EHDS. We therefore analyse in the next section the provisions in the proposed EHDS Regulation and whether they can help overcome current legal

48 Finland, Data Protection Act 1050/2018 (n 42); Luxembourg, Loi du 1er août 2018 portant organization de la Commission nationale pour la protection des données et mise en oeuvre du règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données), portant modification du Code du travail et de la loi modifiée du 25 mars 2015 fixant le régime des traitements et les conditions et modalités d'avancement des fonctionnaires de l'État <https://legilux.public.lu/eli/etat/leg/loi/2018/08/01/a686/jo> (accessed 4 September 2024).

49 Lag (2003:460) om etikprövning av forskning som avser människor 2003.

50 Tzortzatou-Nanopoulou and others (n 43).

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

hurdles in secondary use of data in Europe, at least insofar as electronic health data are concerned.

Why secondary use in the context of the proposed EHDS Regulation remains problematic

The theoretical analysis presented in 'Challenges with the existing GDPR legal bases for processing health and genetic data for secondary use' of this article complements previous exploratory research by others who have reviewed the landscape of the GDPR legal bases across European countries, finding significant heterogeneity in Articles 6(1) and 9(2) GDPR legal bases on which controllers rely to process existing personal health data.[51] Our analysis shows that this heterogeneity, when considered in the context of cross-border secondary use of health and genetic data, will inevitably result in situations where not all controllers across the data reuse lifecycle have a valid legal basis for the processing they are pursuing. This, in turn, is a major hindrance to widespread sharing and availability of such data for secondary uses, in particular where data should be optimized for secondary use across the EU.

Given the challenges associated with the data subject's consent, additional legislative measures are necessary to ensure the availability of a valid GDPR legal basis for all the controllers involved in the data reuse lifecycle. In principle, legislative developments can take place at both the national and the EU levels. However, from a pragmatic point of view, it is the former approach, ie, EU-level legislation, that seems viable. The national implementations of the GDPR have already diverged so dramatically in their interpretation of GDPR Articles 6(1) and 9(2)[52] that any attempts at reversing this divergence in favour of harmonized implementation are, in our view, unlikely to succeed. Therefore, a more pragmatic path towards ensuring the availability of valid GDPR legal bases for cross-border secondary use of health data is through EU legislation. Such thoughts were also reflected in the Joint Action TEHDAS report, 'Why health is a special case for data governance'.[53]

The desire for a pragmatic path (as evidenced by the impact assessment performed by the Commission[54]) was part of the main motivation behind the proposed EHDS Regulation, whose draft proposal was published in May 2022 by the European Commission.[55] One of the main aims of the proposed Regulation is to overcome the legal hurdles for secondary use of health and related data stemming from the differences in the implementation and interpretation of the GDPR in the Member States. This includes, among other things, addressing the insufficient availability of Articles 6(1) and 9(2) GDPR legal bases for the data processing operations across the data reuse lifecycle.

The proposed EHDS Regulation: how does it create a legal basis?

Through the proposed EHDS Regulation, the EU has set out to improve access to, and secondary use of, 'electronic health data', intended to encompass all forms of data related to health as well as administrative healthcare data. In the context of the secondary use of electronic health data, the most relevant legal roles defined by the proposed EHDS Regulation, based on the most recent compromise text published at the time of writing,[56] are those of the Health Data Holder, the Health Data Access Body (HDAB), and the Health Data User. These three EHDS roles loosely correspond,[57] respectively, to the three terms used throughout this article: the Data Provider, the DSI (where the DSI is a permit authority and can mandate the Data Provider to share the data), and the Data User. In this section, we use the EHDS terminology.

The following legal framework is established for secondary use, as per recital 37 of the proposed EHDS Regulation:

- Health Data Holders act under a legal obligation (Article 6(1)(c) GDPR combined with a suitable

51 Hansen and others (n 15).

52 ibid. See also Roxanne Meilak Borg and Mireille Martine Caruana, 'Alternative Legal Bases for Processing Health Data for Scientific Research Purposes' (2024) 18 Masaryk University Journal of Law and Technology 3.

53 Catia Pinto and others, 'Why Health is a Special Case for Data Governance' (TEHDAS Milestone 5.7, 23 June 2021) <https://tehdas.eu/tehdas1/app/uploads/2021/06/tehdas-why-health-is-a-special-case-for-data-governance-2021-06-23.pdf> (accessed 4 September 2024).

54 European Commission, 'Commission Staff Working Document: Impact Assessment Report Accompanying the Document Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space' (Strasbourg, 3 May 2022, SWD (2022) 131 final) <https://health.ec.europa.eu/document/download/6d73ef73-6480-443b-a796-34b4bee3c5ee_en?filename=ehealth_ehds_2022ia_1_en_0.pdf> (accessed 4 September 2024).

55 European Parliament and The Council, Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space 2022 [2022/0140].

56 We refer here to the compromise text of the EHDS Regulation proposal adopted in April 2024. See European Parliament legislative resolution of 24 April 2024 on the proposal for a regulation of the European Parliament and of the Council on the European Health Data Space [COM (2022)0197—C9-0167/2022–2022/0140(COD)] <https://www.europarl.europa.eu/doceo/document/TA-9-2024-0331_EN.html> (accessed 4 September 2024).

57 For a detailed comparison of the terms, see Table A1 in the Annex.

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

exemption under Article 9(2) GDPR) where they process personal electronic health data for the characterization of the data and when they disclose relevant datasets to the HDAB following the issuance of a data permit.
- HDABs act under a task in the public interest (Article 6(1)(e) GDPR combined with a suitable exemption under Article 9(2) GDPR) when they prepare and disclose data in the context of a data access request.
- Health Data Users are required to demonstrate their own legal basis under either Article 6(1)(e) GDPR or Article 6(1)(f) GDPR based on their own situation, but the Article 9(2) GDPR exemption to process health and genetic data is provided through the legal framework of the EHDS.

We note that the creation of GDPR legal bases by the EHDS Regulation proposal, as per recital 37 in the proposed EHDS Regulation, is incomplete in this respect, failing to fully address all elements relevant for an efficient cross-border use of health and genetic data.

The proposed EHDS Regulation and GDPR legal bases: unresolved challenges

On the face of it, the proposed EHDS Regulation appears to resolve the issues with the unavailability of GDPR legal bases, making our analysis in the first half of this article moot. More specifically, it appears to ensure that all three controllers are provided with Articles 6(1) and/or 9(2) GDPR legal bases in a standardized manner. Even in the exceptional case of the Health Data User, for which the proposed EHDS Regulation does not create an Article 6(1) GDPR legal basis, the overall impact is undoubtedly enabling through the creation of Article 9(2)(g–j) GDPR exemptions alone. As we discussed previously in the above section 'The challenge with legitimate interest: GDPR Article 6(1)(f)', many (Health) Data Users will be able to rely on Article 6(1)(f) GDPR, provided they meet the criteria such as performing a balancing test and demonstrating real and present legitimate interest. As the processing of personal data in accordance with the Article 6(1)(f) GDPR does not require recourse to another law, the main practical challenge with this legal basis has been the unavailability of a corresponding Article 9(2) GDPR exemption, which is precisely the barrier to be removed by the EHDS Regulation.

However, a closer look reveals that this purported creation of the GDPR legal bases across the entire health data reuse lifecycle and for all three controllers is deceptive. The reality is that the legislator may have failed to account for the different phrases in the data reuse lifecycle in a sufficiently granular manner, resulting in missing or insufficient GDPR legal bases for undertaking certain crucial processing operations.

A notable example in this respect is that there is no GDPR legal basis provided by the proposed EHDS Regulation to enable Health Data Holders to proactively prepare value-added, reuse-ready electronic health datasets by, for example, cleaning, standardizing, and harmonizing datasets into defined data models. Rather, when it comes to making electronic health data available through the EHDS ecosystem, it appears that, independent of a specific data access request, the only type of processing covered under the GDPR legal basis that is created by the EHDS Regulation is limited to findability by creating dataset descriptions for the data catalogue. No legal basis is created to hold data for secondary use or to curate them into defined data models and standards in a generic manner, unless and to the extent this processing is necessary for the specific purpose pursued by the Health Data User. This affects tremendously the usefulness of data as they were optimized for primary rather than secondary use purposes. A lack of harmonization also impedes findability of relevant subsets of data, which will be impossible to single out across non-harmonized data. Last but not least, a repeated extraction of data for each request will also put a strain on the Health Data Holder.

The issue of the lack of a legal basis to hold data for secondary use is also mirrored at the end of the Health Data User, which also lacks GDPR legal bases for the personal data processing operations beyond those necessary for the accomplishment of the Health Data User's purpose. This means that, for example, when enriched data are created during the processing required by the Health Data User, such as through data curation, standardization, or annotation, the Health Data User has no valid legal basis to have the resultant enriched dataset preserved for own or other parties' future uses. Nor is it clear to what extent they can lawfully act as controllers for the purpose of retaining this enriched data. It appears the only possibility to lawfully store these data is if the Health Data Holder deems the enriched data suitable for their purposes. As there is no legal basis envisaged for storing the data beyond the processing for the Health Data Holder's own purposes, it follows from the GDPR that the potentially valuable enriched data may have to be deleted, unless the Health Data Holder can demonstrate the enriched data are fit for its own purposes (as per Article 41 of the proposed EHDS Regulation) or national legislation has been set up. While this option is explicitly provided for in

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

paragraph 8a of Article 33 in the compromise text of the EHDS Regulation, it is not guaranteed that Member States will enable this option. This will, in particular, hamper the possibility of holding on to improved data from cross-border access requests using data from several countries. Even though the Health Data User can request the storing of the formula on the creation of the requested dataset, the evolvement of datasets over time will likely render this alternative option ineffective.

The EHDS further failed to achieve one of its very goals, which was *establishing a common mechanism to access electronic health data for secondary use across the Union* (recital 37 EHDS compromise text). Paragraph 5 of Article 33 reintroduces the possibility of Member States having their own rules when it comes to processing genetic and genomic data (among others). This option is not limited to the conditions under which data are included in the EHDS, so in consequence, Member States are even able to introduce additional requirements for Health Data Users. This is further widened to all personal electronic health data as Article 45(2)(ha), as well as Article 46(1)(e) and (g), foresee the option that national ethics legislation has to be complied with as well, except where consent is required as a legal basis (see recital 37).

While cross-border harmonization of data for secondary use directly enabled through the proposed EHDS Regulation is not foreseen, pan-European harmonized data will likely be found in the cross-border registries or databases of electronic health data described in Article 53. The harmonization has taken place outside the EHDS when the respective Health Data Holder(s) have assembled the database or registry for their own purposes and Article 53 aims to make these data available in a single mechanism. The proposed solution for dealing with these cross-border data collections seems logical where the database has been assembled by a single controller. However, the situation may become more delicate where joint controllership of such a database is applicable, which would, for example, be the case where data are brought into a joint research project by different organizations.[58] In such a case, data collected by an organization in country A and jointly used by researchers in country A and B would, through this mechanism, make the data also available through the HDAB in country B. Personal health data associated with entire patient cohorts that were established in one country may through this mechanism come under the administration of an HDAB in another country. The same would be true of a single network of registries or databases established at the Union level, such as, for example, the European Reference Networks (ERNs) where healthcare professionals bring in data of rare disease patients to find other healthcare professionals for advice and second opinions.[59] In such a situation, the HDAB in the country of the coordinator of the network effectively becomes the permit authority instructing all participating controllers of registries in this network, irrespective of where they are based. It remains to be seen if this pragmatic approach will be acceptable to the EU Member States and/or the Health Data Holders, or whether it will diminish the willingness for cooperation, in particular in view of the observation above that no full agreement could be achieved among the Member States on the common mechanisms.

Collectively, these considerations point to a problematic conclusion that the proposed EHDS Regulation, in its current form, effectively fails to create a GDPR legal basis for the purposes of generating standardized datasets that can be processed for secondary use *across the entire EHDS ecosystem* in a consistent manner. This reality runs the risk of undermining the utility of the electronic health data undergoing processing in the EHDS and the charm of unified data governance for secondary use throughout the EU.

Last but not least, while the proposed EHDS Regulation overcomes the legal limitations for Health Data Users, whose ability to lawfully use the data is currently hampered due to limiting Article 9(2) GDPR implementations or specific requirements based on Article 9(4) GDPR, this benefit applies exclusively where the permission to reuse the data is granted by the HDAB. Health Data Users who want to access electronic health data from other sources, such as dedicated repositories specialized in enabling secondary use based on their own legal basis outside the EHDS, may discover that their specific use case is not covered in the scenarios for which the EHDS Regulation seeks to create a legal basis.

The proposed EHDS Regulation—conclusion regarding its utility for harmonized secondary use in a cross-border EU-wide data space

Although the proposed EHDS Regulation is an important step in the direction of creating a valid GDPR legal basis, our analysis shows that the EHDS Regulation, in its current form, is inadequate in three substantive areas.

58 EDPB, Guidelines 07/2020 (n 20), 22.

59 European Commission, 'European Reference Networks' <https://health.ec.europa.eu/rare-diseases-and-european-reference-networks/european-reference-networks_en> (accessed 4 September 2024).

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

First, it does not provide a legal basis to improve and hold data for secondary use. Data are held by Health Data Users for their own purposes. If these purposes cease, data are (expected to be) deleted. There is no legal basis to improve data for secondary use as data are—according to EHDS's rules—extracted directly from the primary use context every time and only if there is a defined access request (as per the duties of Health Data Holders under proposed EHDS Regulation Article 41). Any improvements of data for secondary use will therefore need separate legislation, where it needs to be ensured that such legislation does not let the data fall outside the EHDS (if the entity subsequently holding the data is no longer meeting the definition of a Health Data Holder). This is also demonstrated by the proposed EHDS Regulation in Article 33(8a) of the compromise text, as the Health Data Holder may not have a legal basis to retain improved data and to avoid data being lost, and Member States can provide through legislation a legal basis to keep the improved data. In consequence, data remain with their Health Data Holders, are not harmonized, and must be extracted again and again for each data access request unless there is additional (currently still missing) legislation established.

Secondly, it does not promote fully harmonized data access and use criteria: Article 33(5) of the proposed EHDS Regulation enables Member States to introduce stricter safeguards beyond the EHDS requirements for genomic and other molecular data as well as data from wellness apps and biobanks with their associated databases in general. In addition, national ethics legislation can introduce further requirements to be fulfilled in data access requests, as per Article 45(2)(ha) as well as 46(1)(e) and (g) of the compromise text.

Finally, it does not promote easy cross-border access to data. Each HDAB decides on data of the respective Member State. That means for cross-border access, a Health Data User will have to undergo a review in each country and depend on the extraction of data from all the relevant Health Data Holders in the respective countries. While Article 54 addresses the possibility of mutual recognition, paragraph 3a in Article 46 of the Regulation clearly states that each HDAB will be responsible for the access request decision in case of a cross-border access request. Accountability will therefore require each HDAB to perform its own analysis of the health data access request. An effective transfer of responsibility is in principle established through Article 53 of the proposed EHDS Regulation where a single HDAB becomes responsible for the access decisions on data from joint controllers or even networks of independent controllers. However, in view of different practices and ethical concepts across the Member States, this provision may even negatively affect research as it could discourage collaboration in order to not lose national control over the data.

In sum, then, additional legislative measures may be required if electronic personal health data are to be processed in a standardized manner through the EHDS ecosystem, and it is to these possible measures that we now turn.

Some proposed legislative solutions

It is useful to recap the principal setup for secondary use. For each of the steps to make data available for secondary use we discussed above—(i) to make data in principle available for secondary use; (ii) to disclose data to a Data User for their particular purpose; and (iii) for the Data User to pursue their purpose(s)—an independent legal basis under the GDPR has to be established. Indeed, the controller for each of these purposes can be different, which is the case where a DSI takes over the downstream data access decisions.

DSIs may be established under national laws, where the legislation normally foresees both the legal basis for the initial controller who collected the data to transfer the data to a defined DSI, and for the DSI to retain and subsequently disclose the data to Data Users for defined secondary use purposes (eg, the Danish NGC). These purposes are currently mostly limited to research, thus precluding a wider use of these data for other purposes such as healthcare reuse or policy development. As these data were collected based on legislation, the DSI has no legal basis to disclose the data for other purposes not included in the legal mandate.[60] Also, the situation of the Data Providers is difficult: they may have a legal obligation to transfer data to a specific DSI under certain circumstances and/or upon an explicit request by the DSI, but there are few possibilities that would allow them to do so at their own initiative. For example, to proactively make data available through a DSI, Data Providers will not be able to rely on the Article 6(1)(e) legal basis—performance of a task in the public interest—unless making data available for secondary use is explicitly foreseen as part of their mission prescribed by the relevant law or another law applicable to them. Open science policies may give rise to the expectation that the data generated through publicly funded research must be made widely available for reuse, but under the current EU data protection framework this is not viable. The necessity to explicitly address open science requirements in organizational missions of public sector research bodies does not yet seem to be widely recognized.

60 Becker and others (n 3).

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

The open science policies and requirements laid down in the legislative acts that establish public research funding programmes could be considered as a possible source of a GDPR legal basis to make data generated through such programmes widely available for secondary uses. Yet, typically the legislation is too generic and/or not directly applicable to researchers. An example is the EU funding programme Horizon Europe (HE). In Article 1 of the Regulation establishing the programme, it is stated that the Regulation 'sets out the rules for participation and dissemination concerning indirect actions under the Programme and determines the framework governing Union support for R&I activities for the same duration'.[61] This indicates that the Regulation is limited in scope to the programme itself, not the overall governance of the entities that receive funding. Accordingly, in Article 14 HE, on open science, the Regulation merely foresees that open science approaches should be 'encouraged' and 'promoted'. These provisions are not sufficient to create a legal basis under the GDPR for data reuse as no clear task is assigned to the funded organizations themselves. But even if research funding programmes were to be more explicit regarding establishing a valid legal basis for secondary use, such secondary uses would likely be limited to research purposes due to the context and the nature under which the funding programme was set up.

As already discussed above, the proposed EHDS Regulation in its current form will also not provide a legal framework to improve the systematic availability of the data that are not just findable, but also interoperable. To make up for this limitation of the EHDS Regulation, there may be an opportunity forthcoming due to the implementation of the EU's Data Governance Act by the EU Member States. This Act requires the Member States to implement certain features and bodies for secondary use, which means that national legislators across all Member States will be required to revisit their existing legal frameworks concerning secondary data use. This could be an opportunity to remedy the situation and take measures to improve the general availability of data.

In the current setup, then, this leaves the data subject's (explicit) consent, that is, Article 6(1)(a) and Article 9(2)(a), as the most likely choice of legal basis. A valid consent could be obtained to make data available for secondary use where different categories of purposes are sufficiently delineated and can be consented separately. It is also required that the data governance under which data are made available be known, alongside the categories of future recipients.

Yet, in addition to the limitations we already noted above, consent reaches its limits as a feasible legal basis when the downstream data disclosure arises. Specificity of consent means that data subjects would have to consent to *every instance* of data disclosure for a new specific purpose, or to a new data controller (provided that the new controller intends to keep processing the data based on the data subject's consent as the GDPR legal basis). There are ongoing discussions at the EU level regarding the extent to which a broad consent to the future use could also cover the downstream data disclosure. A statement by the German '97. Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder' seems to be more receptive to the notion of a broad consent: while at first an approach is pursued that requires a partial identification of the purpose as prerequisite for a broad consent, under the discussion of safeguards, further research purposes ('weitere Forschungszwecke') and new research questions are referred to, thus suggesting that broad consent may be applicable for further processing.[62] On the other hand, the Opinion 110 by the Italian Data Protection Authority (Garante per la protezione dei dati personali) stated that it is still required to obtain a new consent for downstream research projects that have not been fully identified at the time when a research database is established, in line with our considerations above.[63]

As such, our view is that there is a high degree of legal risk in relying on consent for the disclosure of data to downstream research users where this is not substantiated in the law. Where consent is intended as a legal basis to operate a data infrastructure for secondary use in research, it is strongly recommended to perform a DPIA, followed by a subsequent prior consultation with the relevant Supervisory Authority. This can provide more clarification if consent is likely to be accepted as a valid legal basis for the research in the country. However, there is still a remaining risk of a broad consent not being accepted as a valid legal basis under the GDPR were it to be challenged in a court of law. In any case, where the secondary use goes beyond research,

61 European Parliament and the Council, Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe—the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013 (Text with EEA relevance) 2021 [2021/695].

62 'Beschluss der 97. Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder zu Auslegung des Begriffs "bestimmte Bereiche wissenschaftlicher Forschung" im Erwägungsgrund 33 der DS-GVO 3. April 2019' (2019) <https://www.datenschutzkonferenz-online.de/media/dskb/20190405_auslegung_bestimmte_bereiche_wiss_forschung.pdf> (accessed 4 September 2024).

63 Garante per la Protezione dei dati Personali, 'Parere Ai Sensi Del Ai Sensi Dell'art. 110 Del Codice e Dell'art. 36 Del Regolamento' (2022) 9791886.

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

there will be no doubt that a consent must be obtained for each purpose for which data are disclosed to Data Users.

So, what other options are possible, in light of our pessimistic view above?

We suggest that a path worth considering is to create data infrastructures through the legal act that establishes research infrastructures with physical and/or digital research-enabling capabilities. Many pan-European infrastructures have already been established as European Research Infrastructure Consortia (ERICs) *via* the Council Regulation (EC) No 723/2009.[64] The rule that the legislation creating a public body should also provide the legal basis for its mission and related data processing may also apply to ERICs, which are established through a Commission Implementing Decision to which the statutes of the ERIC describing its mission are annexed. However, ERICs are limited to a research mission. The EU legislator seems to have realized that legal instruments are needed that cover a broader scope of purposes. In the Decision (EU) 2022/2481 of the European Parliament and of the Council of 14 December 2022 establishing the Digital Decade Policy Programme 2030,[65] the new legal instrument of a European Digital Infrastructure Consortium (EDIC) is established, exhibiting characteristics similar to the setup of the ERICs. This new legal instrument may allow the creation of DSIs with a mission to make cross-border data collections available under a common data governance, and for a broad range of purposes. Data can be included in such an EDIC either through consent or through national legislation. The latter does actually not have to be harmonized as it merely covers the data governance for data inclusion into the EDIC, which can differ as long as the EDIC itself pursues a harmonized data governance vis-à-vis the Data User.

EDICs making data available for secondary use can become so-called 'authorised participants' in the EHDS as described in Article 52(4) of the proposed Regulation. EDICs connected to the EHDS would remain responsible for their own decision-making on access provision to (Health) Data Users as per Article 46(3a) of the provisional compromise text. Also, the definition of a Health Data User reflects that not only HDABs but also authorized participants can approve and authorize data processing for secondary use by Health Data Users. In consequence, the EHDS Regulation could be assumed to provide for Article 9(2) GDPR permission of Health Data Users also when they process data disclosed by an EDIC that has been connected to the EHDS.

This path may be worth investigating to support the secondary use of personal data. In this regard, the ongoing European projects aimed at establishing EDICs in the health domain, such as EUCAIM and the 1+MG Initiative, are likely to generate valuable implementation and governance insights in the years ahead.

For Data Users, recital 42 of the GDPR leaves no doubt that an individual consent will be required to provide them with a valid legal basis under the GDPR if they want to rely on Article 6(1)(a) as a legal basis, combined with Article 9(2)(a), where applicable. The Data Governance Act has recognized this necessity, likely together with the realization that the legal framework in some Member States does not always offer an alternative to consent. However, rather than enabling a GDPR legal basis other than consent, the DGA states that public sector bodies who want to disclose data to users are encouraged to support the users in obtaining consent, as per Article 5(6) of the Act.

Conclusion

Despite the EU's 'European Strategy for Data' pointing out data availability as the primary problem that necessitates creating a European data space,[66] none of the subsequent laws proposed or passed by the EU as yet provide for a fully adequate legal framework that enables the creation of pan-EU data resources supporting effective, responsible secondary use. This is all the more surprising as interoperability is also identified as a problem. However, the EU legislator seems to focus only on the generation of more interoperable data in the primary use rather than considering the transformation of the existing data to streamline its reusability. Such a prospective approach towards building high-quality and interoperable data resources means that considerable time will be required to achieve this goal. At the same time, high-quality, interoperable data generated through secondary use from previously unstructured data are not specifically protected and may be lost at the end of the processing by the user.

This leaves key stakeholders in research and health care with very limited options to make data available for secondary use, much to the detriment of patients and wider society. Consent can be used for principal data availability, at least where there is no significant

64 The Council of the European Union, Council Regulation (EC) No 723/2009 of 25 June 2009 on the Community legal framework for a European Research Infrastructure Consortium (ERIC) 2009 [723/2009].

65 European Parliament and the Council, Decision (EU) 2022/2481 of the European Parliament and of the Council of 14 December 2022 establishing the Digital Decade Policy Programme 2030 (Text with EEA relevance) 2022 [2022/2481].

66 European Commission Communication COM (2020) 66, 'A European Strategy for Data' <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066> (accessed 4 September 2024).

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

imbalance of power between the controller and the data subjects, but in our view and in light of the above analysis, subsequent data sharing with third parties, in order to remain lawful under the GDPR, requires new legislative acts. Research infrastructures established as EDICs may constitute an innovative option to create a common data governance for the disclosure of harmonized cross-border data collections to users.

However, the recent quick succession of Regulations (increasingly formulated as 'Acts') and other legislative actions on the EU level indicates that the field of secondary use is set to remain very dynamic in the years ahead. The proposed EHDS Regulation has undergone changes in the series of 'trilogue' negotiations between the European Commission, Council, and Parliament, but even though a provisional agreement between the three bodies has been reached, the resulting compromise text has still many inconsistencies that make it difficult to interpret the text, and it is unclear to what extent such inconsistencies will be remedied (if at all). More legislation is underway with the EU's recent Data Act,[67] which also includes provisions on 'data spaces'—and additional legislation may be in the pipeline of the EU legislator to implement the 'European strategy for data'. In parallel, the national Member State adaptations and implementation of legislation for the Data Governance Act are being finalized. It remains to be seen if all these newly implemented and proposed laws, individually or collectively, can adequately address the significant GDPR legal challenges impeding the broad availability of valuable data collections for secondary uses. For the sake of patients and wider society, we can only hope positive change, or at least legal clarity, arrives soon.

Funding

Funding support for this article was provided by the European Union's Horizon 2020 research and innovation programme Coordination and Support Action 'Beyond 1 Million Genomes (B1MG)' (Grant Agreement no. 951724), the European Union's Digital Europe Programme project 'Genomic Data Infrastructure (GDI)' (Grant Agreement no. 101081813), and the European Union's EU4Health programme project 'CAN.HEAL—Building the EU Cancer and Health Genomics platform' (Grant Agreement no. 101080009) (co-author R.B.). The work of co-author D.C. was supported by the European Union's Horizon 2020 research and innovation programme Coordination and Support Action Healthy Cloud (Grant Agreement no. 965345) and the Innovative Medicines Initiative Joint Undertaking Research and Innovation Action European Platform for Neurodegenerative Diseases (EPND, Grant Agreement no. 101034344).

Conflict of interest statement

The corresponding author (Prof Edward S. Dove) is an Editor of *International Data Privacy Law*.

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

67 Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonized rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act), OJ L, 2023/2854 <https://eur-lex.europa.eu/eli/reg/2023/2854/oj> (accessed 4 September 2024).

Annex

Table A1. Disambiguation of key terms.

Terms used in this article and their definitions	Related legal terms under EU law, their source, and definition	Comparison
Data Provider: an entity that originally collected personal data for its own purposes and that subsequently wants to make the data systematically available for secondary use to Data Users	**Health Data Holder,** according to Article 2(2)(y) of the proposed EHDS Regulation, is a party in the health care or care sectors or pursuing research in relation to the health care or care sectors that has either: (a) The right or obligation to process personal electronic health data for a set of defined purposes in its capacity as a controller or joint controller; or (b) The ability to make available non-personal electronic health data	On the national level, the EHDS term Health Data Holder is limited to entities with a role in (health)care or research and development for the benefit of health care. This does not include entities whose sole mission is to make data available for secondary use. The phrasing may also limit entities much more to health care and the inclusion of organizations pursuing fundamental research in biomedicine could arguably be excluded. In contrast, our term Data Provider is much more open and flexible, even though for this article, we focus on the likely candidates of biomedical research stakeholders, hospitals and, in some examples, genome centres. On the other hand, actors that only make available anonymous or anonymized health data are also Health Data Holders, an element that is beyond the scope of our analysis this article.
	Data Holder, defined in Article 2(8) of the Data Governance Act (DGA), is a party that 'in accordance with applicable Union or national law, has the right to grant access to or share certain personal data or non-personal data'	The definition of the DGA term Data Holder implies that in the case of personal health data, the Data Holder would have a valid GDPR legal basis to share the personal health data. This is in contrast to the legal reality facing some of the Data Providers, as elucidated in our article.
Data-Sharing Intermediary (DSI) is any actor interposed between the Data Provider and the Data User in the data reuse lifecycle that enables data sharing in a systematic manner, either acting as a processor or as a controller for the disclosure of the data	**Health Data Access Body (HDAB)** has an important role under the proposed EHDS Regulation. Although an explicit definition is lacking, the term can be defined indirectly, based on EHDS Regulation Articles 2(2)(aa), 36, and 46, as an authority issuing data permits for certain electronic health data for the purposes and under the governance rules prescribed in the EHDS Regulation	DSI is a broader term that encompasses HDABs. DSIs additionally include any entity that supports the availability of health and genomic data for secondary use. Moreover, whereas DSIs can either be controllers or processors for granting data access to Data Users, the HDAB acts as the controller when issuing a data permit (Article 51(1) EHDS) and is limited to the EHDS context only.

Continued

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

Table A1. *Continued*

Terms used in this article and their definitions	Related legal terms under EU law, their source, and definition	Comparison
	Health Data Intermediation Entity under the proposed EHDS Regulation can be mandated based on Member State law (Article 32(a) EHDS) to fulfil the tasks of Health Data Holders (Article 32a)	Given the prerequisite that Health Data Intermediation Entities must be established based on national law, they can only constitute a subset of DSIs as defined in the present article. One could even argue that they are not covered by our definition of a DSI at all as they act between the Health Data Holder and the HDAB. They are therefore not involved in the actual data disclosure to the Health Data User as foreseen for the DSIs that we have defined.
	Under the **Data Governance Act** Article 2(11), a **Data Intermediation Service** is defined as 'a service which aims to establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other means, including for the purpose of exercising the rights of data subjects in relation to personal data …'	Under certain circumstances, a DSI in the sense of the present article can also be a Data Intermediation Service provider within the meaning of the DGA. However, this is limited to those DSIs that act as processors engaged by Data Providers in their capacity as controllers.
Data User: a party aiming to use, for its own purpose, the personal health data made available by the Data Provider	A **Data User** under the Data Governance Act is a party that has '… the right, including under Regulation (EU) 2016/679 in the case of personal data, to use that data for commercial or non-commercial purposes' (Article 2 (9) DGA)	An important distinction between the two concepts is that the DGA definition is strictly legal: it presupposes a valid GDPR legal basis to use personal data. By contrast, Data User in the sense of the present article is an operational role: it refers to the fact that the party intends to use personal data for its own purpose, irrespective of whether such use is objectively lawful under the GDPR.

Continued

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

Table A1. *Continued*

Terms used in this article and their definitions	Related legal terms under EU law, their source, and definition	Comparison
	Health Data User, under proposed EHDS Regulation Article 2(2)(z), is a party that 'has been granted lawful access to electronic health data for secondary use pursuant to a data permit, data request or an access approval by an authorised participant in HealthData@EU'	The proposed EHDS Regulation definition of the Health Data User is highly specific to the EHDS context.

https://doi.org/10.1093/idpl/ipae014
Advance Access Publication 4 October 2024

Downloaded from https://academic.oup.com/idpl/article/14/3/223/7811298 by Maynooth University user on 27 May 2025

© The Author(s) 2024. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
International Data Privacy Law, 2024, 14, 223–246
https://doi.org/10.1093/idpl/ipae014
Article