
EUROPEAN JOURNAL OF HEALTH LAW 30 (2023) 129–157
BRILL | NIJHOFF
European Journal of Health Law
brill.com/ejhl

Secondary Use of Personal Health Data: When Is It "Further Processing" Under the GDPR, and What Are the Implications for Data Controllers?

Regina Becker | ORCID: 0000-0002-6711-8375
ELIXIR-Luxembourg, Luxembourg Centre for Systems Biomedicine, University of Luxembourg. Esch-sur-Alzette 4367, Luxembourg
Corresponding author, e-mail: regina.becker@uni.lu

Davit Chokoshvili | ORCID: 0000-0002-7537-1987
Bioinformatics Core, Luxembourg Centre for Systems Biomedicine, University of Luxembourg. Esch-sur-Alzette 4367, Luxembourg
davit.chokoshvili@uni.lu

Giovanni Comandé | ORCID: 0000-0003-2012-7415
Sant'Anna School of Advanced Studies, Piazza Martiri della Libertà 33, 56127 Pisa, Italy
giovanni.comande@santannapisa.it

Edward S. Dove | ORCID: 0000-0003-4095-3677
School of Law, University of Edinburgh. South Bridge, Edinburgh EH8 9YL, Scotland, UK
edward.dove@ed.ac.uk

Alison Hall | ORCID: 0000-0001-6605-4675
PHG Foundation, University of Cambridge, 2 Worts Causeway, Cambridge CB1 8RN, UK
alison.hall@phgfoundation.org

Colin Mitchell | ORCID: 0000-0001-9374-6757
PHG Foundation, University of Cambridge, 2 Worts Causeway, Cambridge CB1 8RN, UK
colin.mitchell@phgfoundation.org

Published with license by Koninklijke Brill NV | DOI:10.1163/15718093-BJA10094
 | ISSN: 0929-0273 (print) 1571-8093 (online)

Fruzsina Molnár-Gábor | ORCID: 0000-0002-9406-2776
Faculty of Law, BioQuant Centre, Heidelberg University,
Im Neuenheimer Feld 267, 69120 Heidelberg, Germany
fruzsina.molnar-gabor@uni-heidelberg.de

Pilar Nicolàs | ORCID: 0000-0003-4166-5941
Faculty of Law, University of the Basque Country, Barrio Sarriena, s/n,
48940 Leioa, Spain
mariapilar.nicolas@ehu.eus

Sini Tervo | ORCID: 0000-0003-3543-3378
Ministry of Social Affairs and Health, Meritullinkatu 8,
00170 Helsinki, Finland
sini.tervo@stm.fi

Adrian Thorogood | ORCID: 0000-0001-5078-8164
Bioinformatics Core, Luxembourg Centre for Systems Biomedicine,
University of Luxembourg. Esch-sur-Alzette 4367, Luxembourg
adrian.thorogood@uni.lu

Abstract

Contemporary biomedical research heavily relies on secondary use of personal health data that were obtained in a different clinical or research setting. Under the European Union's General Data Protection Regulation (GDPR), data controllers processing personal data must comply with the principle of purpose limitation, which restricts further processing of personal data beyond the purpose for which the data were initially collected. However, "further processing" is not explicitly defined, resulting in considerable interpretive ambiguities as to whether "secondary use" of data by researchers constitutes "further processing" under the GDPR. This ambiguity is problematic as it exposes researchers to potential non-compliance risks. In this article, we analyse the term "further processing" within the meaning of the GDPR, elucidate important aspects in which it differs from "secondary use", and discuss the implications for data controllers' GDPR compliance obligations. Subsequently, we contextualise this analysis within a broader discussion of regulating scientific research under the GDPR.

Keywords

GDPR – further processing – secondary use – personal data – biomedical research – scientific research

1 Introduction

Recent years have been characterised by a convergence of rapid progress in medical research, transformative developments in big data analytics, and improved database management methods.[1] Collectively, these trends have been responsible for the growing utility of personal health data obtained from patients, biomedical research participants, and other individuals undergoing medical evaluation. Personal health data are now routinely collected from individuals across a wide range of medical and research settings, and analysed in an integrative manner to accelerate biomedical research, as well as to enable improved prevention, diagnosis, and personalised treatment of diseases.[2,3] The learning health systems and research infrastructures underpinning these developments often process rich, multimodal health data integrated from many sources, which may include electronic health records, clinical notes, medical imaging data, and genomic data, among other data modalities.[4,5]

There is growing recognition that health data generated in different contexts are of significant long-term and multifaceted utility, well beyond the specific clinical, diagnostic, or research use for which they were initially collected.

1 J. Andreu-Perez, C. Poon, R. Merrifield, S. Wong and G.Z. Yang, 'Big Data for Health', *IEEE Journal of Biomedical and Health Informatics* 19(4) (2015) 1193–1208. DOI: 10.1109/JBHI.2015.2450362.

2 S.F. Rose, K. Contrepois, K.J. Moneghetti, W. Zhou, T. Mishra, S. Mataraso, O. Dagan-Rosenfeld, AB. Ganz, J. Dunn, D. Hornburg, S. Rego, D. Perelman, S. Ahadi, MR. Sailani, Y. Zhou, SR. Leopold, J. Chen, M. Ashland, J.W. Christle, M. Avina, P. Limaoco, C. Ruiz, M. Tan, AJ. Butte, GM. Weinstock, GM. Slavich, E. Sodergren, TL. McLaughlin, F. Haddad and M.P. Snyder, 'A Longitudinal Big Data Approach for Precision Health', *Nature Medicine* 25 (5) (2019) 792–804. DOI: 10.1038/s41591-019-0414-6.

3 M.R. Mathis, T.Z. Dubovoy, M.D. Caldwell and M.C. Engoren, 'Making Sense of Big Data to Improve Perioperative Care: Learning Health Systems and the Multicenter Perioperative Outcomes Group', *Journal of Cardiothoracic and Vascular Anesthesia* 34(3) (2020) 582–585. DOI: 10.1053/j.jvca.2019.11.012.

4 L. Lessard, W. Michalowski, M. Fung-Kee-Fung, L. Jones and A. Grudniewicz, 'Architectural Frameworks: Defining the Structures for Implementing Learning Health Systems', *Implementation Science* 12(1) (2017) 78. DOI: 10.1186/s13012-017-0607-7.

5 G. Harerimana, B. Jang, J.W. Kim and H.K. Park, 'Health Big Data Analytics: A Technology Survey', *IEEE Access* 6 (2018) 65661–65678. DOI: 10.1109/ACCESS.2018.2878254.

As a consequence, for well over a decade, professional medical societies and numerous members of the biomedical research community have called for strategies aimed at enabling meaningful secondary use of health data.[6,7,8] In the field of biomedical research, significant efforts have been focused on making health data findable, accessible, interoperable and reusable (FAIR), with the aim of pulling together different types of health data siloed across multiple institutions and storage environments.[9] Governments also support such reuse of data on the national or even European level through initiatives such as the Declaration of Cooperation "Towards access to at least 1 million sequenced genomes in the European Union by 2022" (1+Million Genome Initiative).[10]

Despite significant technological developments towards enabling meaningful secondary use of health data, the overall progress in this area has been hampered by concerns over privacy, confidentiality, and other tangible risks to individuals emanating from routine secondary use of their data.[11,12,13] Health data pertaining to an individual are widely regarded to be of a highly sensitive nature, which necessitates carefully designed safeguards for their responsible

6 C. Safran, M. Bloomrosen, E. Hammond, S. Labkoff, S. Markel-Fox, P. Tang and D. Detmer, 'Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper', *Journal of the American Medical Informatics Association* 14(1) (2007) 1–9.

7 I. Danciu, J.D. Cowan, M. Basford, X. Wang, A. Saip, S. Osgood, J. Shirey-Rice, J. Kirby and P.A. Harris, 'Secondary Use of Clinical Data: The Vanderbilt Approach', *Journal of Biomedical Informatics*, *Special Section: Methods in Clinical Research Informatics* 52 (2014) 28–35. DOI: 10.1016/j.jbi.2014.02.003.

8 S.M. Meystre, C. Lovis, T. Bürkle, G. Tognola, A. Budrionis and C.U. Lehmann, 'Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress', *Yearbook of Medical Informatics* 26(1) (2017) 38–52. DOI: 10.15265/IY-2017-007.

9 P. Holub, F. Kohlmayer, F. Prasser, M.T. Mayrhofer, I. Schlünder, G.M. Martin, S. Casati, L. Koumakis, A. Wutte, L. Kozera, D. Strapagiel, G. Anton, G. Zanetti, O.U. Sezerman, M. Mendy, D. Valík, M. Lavitrano, G. Dagher, K. Zatloukal, G.J.B. van Ommen and J.E. Litton, 'Enhancing Reuse of Data and Biological Material in Medical Research: From FAIR to FAIR-Health', *Biopreservation and Biobanking* 16(2) (2018) 97–105. DOI: 10.1089/bio.2017.0110.

10 *European '1+ Million Genomes' Initiative*, available online at https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes (accessed 26 June 2022).

11 *Supra* note 6.

12 F. Li, X. Zou, P. Liu and J.Y Chen, 'New Threats to Health Data Privacy', *BMC Bioinformatics* 12(12) (2011) S7. DOI: 10.1186/1471-2105-12-S12-S7.

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

secondary use.[14,15,16] Increasingly, these considerations are reflected in various laws and regulatory frameworks, ranging from sector-specific (e.g., laws governing medical secrecy and physician-patient relations) to general (e.g., privacy and general data protection), not to mention various privacy-centred ethical frameworks and guidelines.[17,18]

In Europe (EU/EEA), the General Data Protection Regulation (GDPR), which came into force in 2018, provides a comprehensive regulatory framework governing processing of personal data, including personal health data. The GDPR defines the roles and concrete obligations of parties processing personal data of natural persons.[19,20] These include compliance with the principles of the Regulation, obligations vis-à-vis data subjects, as well as various organisational and technical measures to be implemented. When processing personal health data, which are defined as a special category of data under the GDPR (Article 9(1) GDPR), additional conditions must be satisfied.

However, compliance with the GDPR has been associated with substantial challenges in the context of secondary use of personal data, that is, when the data collected for a particular primary purpose are intended to be used for a different purpose. Healthcare and research institutions collecting personal health data are often confronted with significant uncertainties as to whether, and under what circumstances, they can engage lawfully in secondary use of

14 *Ibid.*

15 O. Choudhury, A. Gkoulalas-Divanis, T. Salonidis, I. Sylla, Y. Park, G. Hsu and A. Das, 'Differential Privacy-Enabled Federated Learning for Sensitive Health Data', *ArXiv:1910.02578* [*Cs*] (2020), available online at http://arxiv.org/abs/1910.02578 (accessed 20 March 2022).

16 M. Jungkunz, A. Köngeter, K. Mehlis, E.C. Winkler and C. Schickhardt, 'Secondary Use of Clinical Data in Data-Gathering, Non-Interventional Research or Learning Activities: Definition, Types, and a Framework for Risk Assessment', *Journal of Medical Internet Research* 23(6) (2021) e26631. DOI: 10.2196/26631.

17 V. Xafis, G.O. Schaefer, M.K. Labude, I. Brassington, A. Ballantyne, H.Y. Lim, W. Lipworth, T. Lysaght, C. Stewart, S. Sun, G.T. Laurie and E.S. Tai, 'An Ethics Framework for Big Data in Health and Research', *Asian Bioethics Review* 11(3) (2019) 227–254. DOI: 10.1007/s41649-019-00099-x.

18 UK Central Digital and Data Office, *Data Ethics Framework*, available online at https://www.gov.uk/government/publications/data-ethics-framework, (accessed 20 March 2022).

19 M. Shabani and L. Marelli, 'Re-Identifiability of Genomic Data and the GDPR: Assessing the Re-identifiability of Genomic Data in Light of the General Data Protection Regulation', *EMBO Reports* 20(6) (2019) e48316. DOI: 10.15252/embr.201948316.

20 T. Bahls, J. Pung, S. Heinemann, J. Hauswaldt, I. Demmer, A. Blumentritt, H. Rau, J. Drepper, P. Wieder, R. Groh, E. Hummers and F. Schlegelmilch, 'Designing and Piloting a Generic Research Architecture and Workflows to Unlock German Primary Care Data for Secondary Use', *Journal of Translational Medicine* 18(1) (2020) 394. DOI: 10.1186/s12967-020-02547-x.

such data in a GDPR-compliant manner. Much of these uncertainties stem from a lack of clear regulatory guidance regarding secondary use of health data, as well as conflicting interpretations of the relevant provisions of the GDPR.[21,22,23] Recent legislative developments, such as the new EU Data Governance Act[24] and the publication of the draft EU Regulation on the European Health Data Space (EHDS),[25] aim to provide a clearer legal framework for the use of data for new purposes. However, it remains to be seen whether these legislative efforts can adequately address the principal GDPR-related challenges associated with the reuse of personal health data.

In this article, we set out to examine secondary use of personal health data through the lens of the GDPR to address and overcome these ongoing uncertainties. In particular, we focus on the closely related GDPR concept of *further processing*, and elucidate under what conditions secondary use constitutes further processing within the meaning of the GDPR and, equivalently, when further processing may not constitute secondary use. Subsequently, we elaborate on the practical compliance implications for pursuing further processing and offer recommendations for ensuring that the various parties involved in data processing are compliant with the GDPR. We conclude by discussing the case of further processing for scientific research purposes, as it is widely regarded that the GDPR affords a special (privileged) status to this type of further processing. We offer novel insights into this matter, emphasising the manoeuvrability of the privileged status of scientific research while also delineating its limits. This

21 E.S. Dove and J. Chen, 'Should Consent for Data Processing Be Privileged in Health Research? A Comparative Legal Analysis', *International Data Privacy Law* 10(2) (2020) 117–131. DOI: 10.1093/idpl/ipz023.

22 A. Cole and A. Towse, 'Data Protection In The European Union Post-General Data Protection Regulation (GDPR): A Barrier Or An Enabler Of Pharmaceutical Innovation?', *International Journal of Technology Assessment in Health Care* 37(S1) (2021) 10–11. DOI: 10.1017/S0266462321000908.

23 A. Vlahou, D. Hallinan, R. Apweiler, A. Argiles, J. Beige, A. Benigni, R. Bischoff, P.C. Black, F. Boehm, J. Céraline, G.P. Chrousos, C. Delles, P. Evenepoel, I. Fridolin, G. Glorieux, A.J. van Gool, I. Heidegger, J.P.A. Ioannidis, J. Jankowski, V. Jankowski, C. Jeronimo, A.M. Kamat, R. Masereeuw, G. Mayer, H. Mischak, A. Ortiz, G. Remuzzi, P. Rossing, J.P. Schanstra, B.J. Schmitz-Dräger, G. Spasovski, J.A. Staessen, D. Stamatialis, P. Stenvinkel, C. Wanner, S.B. Williams, F. Zannad, C. Zoccali and R. Vanholder, 'Data Sharing Under the General Data Protection Regulation', *Hypertension* 77(4) (2021) 1029–1035. DOI: 10.1161/HYPERTENSIONAHA.120.16340.

24 *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)*, available online at http://data.europa.eu/eli/reg/2022/868/oj (accessed 26 June 2022).

25 *Proposal for a Regulation on the European Health Data Space*, available online at https://ec.europa.eu/health/document/download/756d7c59-8641-42a5-94d0-2215f97ec7e5_en (accessed 26 June 2022).

leads us to conclude that data controllers performing scientific research may be subject to tighter obligations under the GDPR than is widely believed.

2 Secondary Use of Personal Health Data

In the scientific literature, there is no broad consensus as to the precise definition of "secondary use" of data. The term is at times used interchangeably with other related terms, including "data reuse" and "repurposing", among others. Conversely, some authors have sought to carefully differentiate these and other semantically related terms, based on factors such as: the degree of contextual similarity between the instances of data uses (i.e., are all instances of data uses taking place in the context of healthcare/research? If so, how similar are the clinical aims and/or research questions across the instances?); whether the existing data are being used by the party that already has the data, or whether a data transfer to another party takes place; and the time interval separating the initial data generation from its subsequent use.[26,27,28,29] While acknowledging the existence of these related terms is important, it is beyond the scope of the present article to engage deeply in their ontological nuance, as our focus rests on the practical implications of secondary use of health data for GDPR compliance. We hope that many of the ambiguities surrounding the definition (or lack thereof) of "secondary use" of health data will be resolved with the advent of the EHDS, whose draft regulation proposal was published by the European Commission in May 2022. The forthcoming EHDS regulation seeks to establish a harmonised European framework governing the lifecycle of personal health data. However, in this article, we broadly refer to "secondary use" of personal health data in a manner that encompasses *any* use of the data beyond the scope for which the data were initially collected or generated. Even so, we acknowledge that this intuitive definition leaves certain ambiguities that are subject to interpretation; the ambiguities which, as we will discuss

26 B. Custers and H. Uršič, 'Big Data and Data Reuse: A Taxonomy of Data Reuse for Balancing Big Data Benefits and Personal Data Protection', *International Data Privacy Law* 6(1) (2016) 4–15. DOI: 10.1093/idpl/ipv028.

27 I.V. Pasquetto, B.M. Randles and C.L. Borgman, 'On the Reuse of Scientific Data', *Data Science Journal* 16(8) (2017) 1–9. DOI: 10.5334/dsj-2017-008.

28 J. Boté and M. Térmens, 'Reusing Data: Technical and Ethical Challenges', *DESIDOC Journal of Library & Information Technology* 39(6) (2019) 329–337. DOI: 10.14429/djlit.39.6.14807.

29 M. Choo and M. Findlay, *Data Reuse and Its Impacts on Digital Labour Platforms* (Rochester, NY: Social Science Research Network, 18 October 2021), available online at https://papers.ssrn.com/abstract=3957004 (accessed 23 March 2022).

later, also account for the fundamental differences between secondary use and further processing under the GDPR.

3 Further Processing of Personal Health Data under the GDPR

The GDPR does not mention "secondary use" of data. The most closely related term in the GDPR is "further processing" of personal data. Although the term is not explicitly defined within the GDPR, based on Recital 50 of the Regulation, it can be deduced that further processing refers to "processing of personal data for purposes other than those for which the personal data were initially collected" (GDPR, Recital 50). Understanding whether a secondary use of data constitutes "further processing" under the GDPR is of major importance as it has a significant impact on the roles and responsibilities of the parties involved in data processing.

First, the principle of purpose limitation under the GDPR generally prohibits data controllers from further processing personal data for a purpose that is incompatible with the purpose that led to the initial data collection. Specifically, Article 5(1)(b) GDPR states that "personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes". Two notable exceptions where this general prohibition of incompatible further processing does not apply are when the further processing is based on: i) the data subject's consent; or ii) a "Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard, in particular, important objectives of general public interest" (GDPR, Article 6(4); Recital 50). In these two cases, further processing is allowed under the GDPR, irrespective of compatibility. Additionally, further processing of personal data for scientific research ("for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes"), when carried out in accordance with Article 89(1) of the Regulation, is considered compatible under the GDPR (Article 5 (1)(b); Recital 50).[30] However, as we will discuss later in this article, there are certain limitations and caveats to applying this presumption of compatibility for research purposes.

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

If these exceptions do not apply (i.e., where further processing is not based on consent or a specific law, and it does not meet the criteria for the presumption of compatibility for scientific research under Article 5(1)(b) GDPR), the data controller must carry out a formal compatibility assessment of the intended further processing activity. This formal assessment process, introduced in Article 6(4) and complemented by Recital 50 of the GDPR, is aimed at ascertaining whether further processing is compatible with the purpose for which data were initially collected. The compatibility test should take into account multiple relevant factors, including the following: "any link between the purposes for which the personal data have been collected and the purposes of the intended further processing" (Article 6(4)(a) GDPR); "the context in which the personal data have been collected" (Article 6(4)(b) GDPR); "the nature of the personal data, in particular whether special categories of personal data are processed" (Article 6(4)(c) GDPR); and "the reasonable expectations of data subjects on the basis of their relationship with the controller as to their further use" (Recital 50 GDPR). Given the multifaceted character of the compatibility test and the considerable judgement required by the data controller, the assessment will need to be carried out on a case-by-case basis.[31]

Moreover, following a positive outcome of the compatibility assessment, the data controller, prior to initiating the intended further processing, may be required to inform the data subject about the intended further processing activity. These requirements are laid down in Arts. 13 and 14 of the GDPR and constitute additional obligations the data controller pursuing further processing of personal data must meet.

In view of the aforementioned requirements, it is critical that medical and research institutions have a clear understanding of whether, and to what extent, these requirements apply to them. In this respect, they need to ascertain whether a particular case of secondary use of personal health data in a given context (e.g., provision of clinical care, or conduct of scientific research) constitutes further processing under the GDPR.

4 When Does Secondary Use of Personal Health Data (Not) Constitute Further Processing under the GDPR?

One of the crucial differences between secondary use in the biomedical research parlance and further processing under the GDPR lies in whether the

31 L. Marelli and G. Testa, 'Scrutinizing the EU General Data Protection Regulation', *Science* 360(6388) (2018) 496–498. DOI: 10.1126/science.aar5419.

processing for an additional purpose concerns an entire data lifecycle, from data generation to deletion (a "data-focussed" view), or a distinct phase within the data lifecycle defined by who is processing the data and for what purposes (a "controller-focussed" view). In the controller-focussed view, each phase within the data lifecycle begins when a controller collects personal data — either directly from the data subject or, in the case of existing data, from another source — and ends with the realisation of the purpose(s) for which that controller collected the data.

Secondary use of personal health data, in the sense of the term used in the present article, implies that the purpose of the subsequent data use differs in a substantial manner from the purpose for which the data were initially generated (i.e., primary use). An example of this would be a hospital that administers medical examinations to an individual in order to diagnose a particular disease, and subsequently decides to use the collected data as part of a research study unrelated to the disease. Given the substantial differences between the initial and subsequent uses of the data, the latter is clearly a secondary use. Importantly, this would also be the case if the research was carried out by a third party to whom the hospital provided the data, as the identity of the entity using the data does not influence the core nature of the data use. To reiterate, secondary use is not a GDPR term and hence this discussion is grounded in how "secondary use" of health data is commonly conceptualised in the biomedical literature.

By contrast, the GDPR does not look onto the entire lifecycle but only onto stages in that lifecycle where the processing is determined by a single or by joint controllers. The GDPR focusses always on the fact that data are processed and why they are processed, building its definitions and framework around the processing operations and the purposes. As an example, controllership is defined based on who determines the purposes and the means of the processing, not based on which data are being processed. It is subsequently this controller who takes the responsibility for all the processing. In line with this approach, further processing under the GDPR is to be understood in relation to the purpose for which a particular controller originally collected the data, whether directly from the data subject or by obtaining existing data from another source. This interpretation is in agreement with the wording in Article 5(1)(b) GDPR, which speaks of data collection in general, as opposed to data collection directly from the data subject.[32]

32 Although the GDPR does not explicitly define "data collection", it can be inferred from different parts of the Regulation that "data collection" is used as a generic term for gathering personal data on the data subject, irrespective of the source of the data. For relevant

With this consideration in mind, we can revisit the example above, where healthcare data are used for biomedical research. Clearly, from a GDPR point of view, further processing does take place since the data were not initially collected for research purposes. However, what may be less self-evident is who performs further processing if more than one party is involved. In other words, where the hospital transfers the patient's data to an external organisation that is the sole party pursuing research, at what point does further processing take place? Based on the controller-focussed approach implicit to the GDPR, we can conclude that it is the hospital engaging in further processing of data by transferring it to the external party. The sharing of data for the recipient's purposes, therefore, constitutes further processing under the GDPR. By contrast, data collection and use for the intended research by the recipient organisation would, under the GDPR, be processing for the recipient's primary purpose, even though under ethical viewpoints such research would be seen as secondary use. This is due to the fact that the data recipient has sought access to the data for the specific purpose of conducting research. By placing an emphasis on the processing activities under each data controller, the GDPR requires that "further processing" be defined in relation to the purpose for which a given data controller has collected the data, as opposed to the purpose for which the data was generated at the start of the data lifecycle.

Another relevant difference between secondary use of data in the biomedical research parlance and further processing under the GDPR stems from the fact that secondary use lacks a precise consensus definition, as noted above, which may give rise to semantic ambiguities. Consider, for example, the case of rare diseases, which is characterised by a convergence of research and

---

parts in the GDPR where "data collection" or "collected data" are used in a generic manner that is agnostic to the source of the data, see, for example, Articles 17(1)(a) and 25(2) of the GDPR. The European Data Protection Board, in its guidance documents, has also implicitly acknowledged this source-independent nature of the term "data collection", having variously discussed "data collected from the data subject" (e.g., The European Data Protection Board (EDPB), *Guidelines 06/2020 on the Interplay of the Second Payment Services Directive and the GDPR | European Data Protection Board* (2020), available online at https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-062020-interplay-second-payment-services_en (accessed 23 March 2022), paragraph 75), and "data collected from third parties" (e.g., The European Data Protection Board (EDPB), *Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies under Regulation 2016/679 | European Data Protection Board* (January 2019), available online at https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-12019-codes-conduct-and-monitoring-bodies-0_en (accessed 23 March 2022), paragraph 16). The terms "collected directly from the data subject" and "obtained" are then used if a difference is to be made between a first generation or collection and a downstream collection of data from another source such as for Articles 13 and 14 GDPR.

clinical diagnosis. Rare disease patients and their family members are commonly subjected to extensive molecular evaluations, including whole exome or whole genome sequencing, which effectively serves a dual purpose: while the evaluation could have an immediate impact on the patient's treatment, the data generated through the evaluation may be retained to enable related research activities, such as statistical analysis of the data from similar patients to identify new clinically significant findings. Any such finding, in turn, could be routinely fed back to the clinical team providing care to the patient, thus potentially impacting on the patient's future clinical management. In this case, it is far less obvious whether, in the common biomedical parlance, the use of the patient data for research purposes would constitute secondary use. Owing to the direct relevance of this research activity to the initial purpose of data collection, the line between primary and secondary use becomes blurry.

By contrast, Article 5(1)(b) GDPR, in the first half-sentence of the subparagraph, stresses that the purpose of data processing must be specific and explicitly defined. This means that under the GDPR, all distinct processing activities, irrespective of their qualitative similarities or relevance to one another, must be clearly delineated with respect to one or more purposes. Viewed in this way, it becomes evident that under the GDPR, subsequent use of the data by a controller that previously collected the data for a different purpose would constitute further processing. This is equally true in the medical contexts where the further processing bears strong relevance to the purpose of initial data collection by the data controller, as with the convergence of medical care and research in the previous example. Of note, the GDPR allows for little flexibility in this regard, as it requires that the purpose of data processing be specific and narrowly defined.[33,34]

33 H.J. Pandit, A. Polleres, B. Bos, R. Brennan, B. Bruegger, F.J. Ekaputra, J.D. Fernández, R.G. Hamed, E. Kiesling, M. Lizar, E. Schlehahn, S. Steyskal and R. Wenning, 'Creating a Vocabulary for Data Privacy', in: H. Panetto, C. Debruyne, M. Hepp, D. Lewis, C.A. Ardagna and R. Meersman (eds), *On the Move to Meaningful Internet Systems: OTM 2019 Conferences* (Cham: Springer International Publishing, 2019) pp. 714–730. DOI: 10.1007/978-3-030-33246-4_44.

34 D. Georgiou and C. Lambrinoudakis, 'GDPR Compliance: Proposed Guidelines for Cloud-Based Health Organizations', in: S. Katsikas, F. Cuppens, N. Cuppens, C. Lambrinoudakis, C. Kalloniatis, J. Mylopoulos, A. Antó, S. Gritzalis, W. Meng and S. Furnell (eds), *Computer Security, Lecture Notes in Computer Science* 11980 (Cham: Springer International Publishing, 2020) pp. 156–169. DOI: 10.1007/978-3-030-64330-0_10.

5 Practical Considerations for the Identification of Further Processing Operations

To help medical and research institutions determine whether a particular processing operation performed on personal health data would constitute further processing under the GDPR, we propose some guiding questions. However, the approach discussed below may be of broader relevance and help other stakeholders accurately identify further processing in contexts beyond medical and research uses of data. As also revealed in a recent court case in Ireland, many stakeholders — including regulatory authorities — may find it challenging to correctly assess whether a processing operation on personal data constitutes further processing.[35] We recommend that in order to distinguish between processing for the primary purpose Y driving the data collection and further processing (for other purpose(s)), institutions ask the following question: "would purpose Y still be achieved in the absence of processing activity P?" We suggest that if the answer to this question is "yes", the processing activity P should be considered processing for a separate purpose, and hence further processing. Importantly, further processing may include processing activities that are closely related to the primary purpose of data collection, as in the example of reporting all diagnostic tests results for an infectious disease described in Table 1 below. Even though reporting of test results is mandated by the law and can therefore be seen as a necessary step integral to the testing process (the primary purpose of data collection), it nevertheless constitutes further processing under the GDPR because data collection would still take place in its absence.

According to Article 5(1)(b) GDPR, personal data must be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes". The use of the plural form (*purposes*) in this sentence clearly implies that under the GDPR, data collection may be driven by more than one primary purpose. However, it is important to highlight that the possibility of having multiple primary purposes cannot be interpreted as permitting categorisation of a purpose as primary at the data controller's discretion. Not all intended purposes known, or even explicitly communicated at the time of data collection, are primary. Some of the predefined purposes can be opportunistic, using the available data that have been collected anyway. Hence, processing the collected data to accomplish these (opportunistic)

35 *The Data Protection Commissioner v Doolin* [2022] IECA 117, available online at https://www.courts.ie/acc/alfresco/6ca61b58-4057-4572-99f5-5a93146d0bb6/2022_IECA_117%20(Unapproved).pdf/pdf (accessed 26 June 2022).

TABLE 1 Examples of Data Processing Activities Involving Personal Health Data Previously Collected by the Data Controller

Processing activity	Further processing?	Secondary use?
A diagnostic laboratory performing testing for an infectious disease routinely reports test results to the national infectious disease registry, in compliance with a national health law.	Yes. Reporting of test results to the registry is further processing under the GDPR. In the absence of this reporting obligation, the laboratory would still collect the data for the primary purpose, i.e., to diagnose the presence of an infection in the tested individual.	Ambiguous Owing to the ambiguity of the term "secondary use", two interpretations are possible: 1. Due to the legal obligation, reporting of test results cannot be disentangled from performing the test. Rather, reporting test results can be seen as an integral step in the testing process, effectively placing it within the scope of primary use. 2. Although the use of data for diagnostic purposes cannot be procedurally disentangled from reporting obligations, the two purposes of patient care and reporting in a wider public health context are sufficiently different to constitute distinct uses of data.
A medical hospital that administered diagnostic evaluation to one of its patients uses the resultant data to help inform diagnosis in another patient treated at the same hospital.	Yes. The hospital generated the data for the purpose of establishing a medical diagnosis in a particular patient. This would have taken place irrespective of the intention to use the data for aiding diagnoses in other cases.	Yes The data are used in different clinical cases and there is no indication that the subsequent use of the data will be directly relevant to the patient whose medical examination lead to the data generation.

TABLE 1 Examples of Data Processing Activities Involving Personal Health Data (*cont.*)

Processing activity	Further processing?	Secondary use?
Medical hospital A shares its records, which it collected from its patients, with medical hospital B. Medical hospital B intends to use the data for aiding diagnoses in its own patients.	Controller-dependent Hospital A: Yes Data disclosure by Hospital A constitutes further processing under the GDPR. Supporting Hospital B's diagnostic operations is not a primary purpose for which Hospital A collected the data. Hospital B: No Obtaining the data by Hospital B and using it for the intended purpose of diagnosing its patients is primary processing.	Yes The data were generated by Hospital A to diagnose particular patients undergoing medical evaluation. Upon generation, they were not intended to be used by Hospital B to facilitate diagnosis in other patients.

purposes would constitute further processing. Such distinction is important where the processing is not based on the data subject's consent or a legislation and may therefore require a compatibility test, as would be the case when data are further processed for quality management purposes, to name one example. An indication of whether a purpose is a parallel primary purpose or an opportunistic purpose can be gained by asking the following question: "would the data collection for purpose X still take place in the absence of purpose Y?" If the answer to this question is "yes", then purpose X is also a primary purpose. If the answer is "no", then X is likely an additional, opportunistic purpose leading to further processing.[36] For example, even if patients upon hospital admission are informed about a possible use of their data for quality assurance or medical research, these processing activities constitute further processing as the collection did not take place for such purposes specifically. Once again, we can use

36 In this example, we assume for this test that both purpose X and Y use the same data and no additional data are collected for purpose X specifically.

the approach recommended above to ascertain that these intended processing activities constitute further processing under the GDPR: we can ask whether, in the absence of intended data use for healthcare (primary purpose), data collection from patients for quality assurance and medical research would still take place. Since the answer to this question is "no" in case of general hospital admission, we should treat these intended processing activities as further processing, carried out to achieve an opportunistic, non-primary purpose.

Table 1 provides several other examples of data processing activities involving personal health data previously obtained by a data controller. The table describes whether the processing activity constitutes further processing under the GDPR or a secondary use in the sense of the term commonly used in the biomedical literature.

5.1 *Implications for Data Controllers: Compatibility of Purposes*

The findings above have several important implications in relation to the roles and legal obligations of data controllers.

First, further processing is always performed by a data controller that has already collected the data for another purpose, as opposed to a controller that is requesting access to the data to process them for its own (primary) purpose. As a consequence, when personal data are shared with an external party and such sharing is not necessary for the purpose for which the data were collected by the controller, this disclosure constitutes further processing. In line with Article 6(4) GDPR, a controller pursuing further processing has to perform a compatibility test, unless: i) the controller can rely on the consent of the data subject for the processing; or ii) the processing is based on a Union or a Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1) GDPR. When either of these two conditions is fulfilled, data can be processed irrespective of the compatibility, as also stated in Recital 50. The background for the special status of these legal bases is that either the data subjects themselves legitimise the processing, and hence render the test redundant, or the balancing of interests and relevant safeguards have been considered in the legislation (as reflected in the reference to the law as a necessary and proportionate measure). In contrast, the required test of compatibility for the transfer includes the compatibility of the purpose for which the recipient controller wants to process the data. Consequently, the question to be answered by the disclosing controller is the following: "Is the sharing of data for the purpose of the recipient compatible with the purpose for which I collected the data?" Thus, the burden of performing the compatibility test falls on the data-transferring controller rather than the data-recipient controller. In

practice, this would mean that the data-transferring controller must carry out the compatibility assessment, where applicable, for the intended further processing activities, and provide the data subject with appropriate information, as mandated by the GDPR principle of transparency.[37]

Second, it is possible for data to be shared with multiple independent controllers without any further processing taking place. In the context of personal health data, this is most notably the case where the data are collected from individuals with an explicit primary purpose of making the data available for external parties. For example, population biobanks often collect and store large amounts of personal health data in their health repositories, alongside biological samples from the data subjects, with the goal of enabling external parties to use these resources in various contexts, such as provision of healthcare to the data subjects, performing medical research, and facilitating health policy development.[38] In the case of such biobanks, transfer of data subjects' personal health data to other parties, such as healthcare providers and biomedical researchers, would not constitute further processing, as the intention to make the data available was the primary purpose driving data collection. At the same time, accessing and using the data by a third party for a specified purpose (e.g., facilitating healthcare delivery, or performing research) would also constitute primary processing of the data. In this scenario, no further processing takes place and, consequently, none of the parties involved in this particular part of the data lifecycle has the obligation to comply with the GDPR requirements for further processing of data. Of note, this conclusion is unaffected by contextual factors such as the duration of data storage and whether data collection from data subjects takes place at a single or multiple time points, the latter being an approach commonly utilised by population biobanks for longitudinal observational cohorts.

6 Further Processing for Scientific Research Purposes

A special case of further data processing that warrants particular attention is further processing for scientific research purposes (that is, further processing "for archiving purposes in the public interest, scientific or historical research

37 Articles 13(3) and Article 14(4) GDPR require controllers to inform the data subject about any further processing they intend to pursue.

38 L. Coppola, A. Cianflone, A.M. Grimaldi, M. Incoronato, P. Bevilacqua, F. Messina, S. Baselice, A. Soricelli, P. Mirabelli and M. Salvatore, 'Biobanking in Health Care: Evolution and Future Directions', *Journal of Translational Medicine* 17(1) (2019) 172. DOI: 10.1186/s12967-019-1922-3.

purposes or statistical purposes"). This type of further processing, when carried out in accordance with Article 89(1) GDPR, "shall [...] not be considered to be incompatible" under Article 5(1)(b) GDPR. Somewhat similarly, Recital 50 of the Regulation states that this form of further processing "should be considered to be compatible lawful processing operations". Although the equivalence of these two statements has been debated, it is clear that they both provide grounds for exempting data controllers undertaking scientific research from the requirements of a compatibility test under Article 6(4) GDPR for further processing.[39] These considerations have contributed to the commonly held opinion that scientific research occupies a privileged status in the GDPR,[40,41] relieving controllers from the compatibility test of Article 6(4) if they are processing personal data for research purposes. Importantly, the aforementioned provisions of the GDPR do not specify that in order for the exemption to apply, the scientific research must be carried out by the data controller that has already obtained the data. Disclosing data for another controller's research project still means processing for research purposes, in this case the research purpose of a third party, as there is no other reason than the research project to share the data. This should be interpreted to mean that sharing of personal data with other controllers who then intend to use the data for their own scientific research purposes must also be a compatible processing operation. However, it should be emphasised that there are some caveats concerning the assumption that further processing of personal data for scientific research purposes is always lawful as suggested by Recital 50 GDPR.

First, the definition of scientific research, in the sense of the intended meaning within the GDPR, is vigorously debated.[42,43,44] In 2020, the European Data

39 G. Schneider and G. Comandé, *Differential Data Protection Regimes in Data-Driven Research: Why the GDPR Is More Research-Friendly Than You Think* (Rochester, NY: Social Science Research Network, 14 July 2021), available online at https://papers.ssrn.com/abstract=3897258 (accessed 23 March 2022).

40 V. Chico, 'The Impact of the General Data Protection Regulation on Health Research', *British Medical Bulletin* 128(1) (2018) 109–118. DOI: 10.1093/bmb/ldy038.

41 P. Quinn, 'Research under the GDPR — a Level Playing Field for Public and Private Sector Research?', *Life Sciences, Society and Policy* 17(1) (2021) 4. DOI: 10.1186/s40504-021-00111-z.

42 N. Clarke, G. Vale, E.P. Reeves, M. Kirwan, D. Smith, M. Farrell, G. Hurl and N.G. McElvaney, 'GDPR: An Impediment to Research?', *Irish Journal of Medical Science* 188(4) (2019) 1129–1135. DOI: 10.1007/s11845-019-01980-2.

43 G. Verhenneman, K. Claes, J.J. Derèze, P. Herijgers, C. Mathieu, F.E. Rademakers, R. Reyda and M. Vanautgaerden, 'How GDPR Enhances Transparency and Fosters Pseudonymisation in Academic Medical Research', *European Journal of Health Law* 27(1) (2020) 35–57. DOI: 10.1163/15718093-12251009.

44 *Supra* note 39.

Protection Supervisor (EDPS) issued a Preliminary Opinion on data protection and scientific research, in which the EDPS argued that "genuine" research needs to conform to the values and principles governing the scientific research framework in the EU. These principles include pursuit of the common good in the public interest, knowledge creation, and reliance on the safeguards commonly utilised in the research context, such as an ethics review process and scientific codes of conduct. The EDPS particularly voiced the concern that non-academic for-profit entities sometimes exercise undue influence on the research to obtain desired outcomes and may therefore not satisfy the requirements for a research exemption.[45] On the other hand, even some of the data processing activities carried out by entities with an unambiguous mandate of doing "genuine" research, such as public universities, may fall beyond the scope of scientific research. This is notably the case when using personal data for quality assurance purposes, which does not aim at improved products or services. In several European countries, the distinction between scientific research and quality assurance activities is even explicitly spelled out in the national law. For example, the Norwegian Health Research Act[46] expressly excludes processing of personal health data for quality assurance purposes from the scope of scientific research.[47]

Second, even where it can be established that a further processing of personal data clearly falls within the scope of scientific research and therefore constitutes a compatible processing operation under Article 5(1)(b) GDPR, this should not be misconstrued to mean that the intended processing is automatically permissible. There are additional GDPR requirements derived from the provisions in Article 5 GDPR alongside, where applicable, legitimations under Article 9 GDPR or safeguards defined in Chapter V GDPR. Moreover, other legal and regulatory constraints, such as those based on the discretion of EU Member States to introduce additional restrictions for processing health and genetic data, could potentially limit the data controller's ability to engage in the intended further processing. As an example, further processing of special categories of data for research purposes requires an approval by the

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

46 The GDPR is a regulation with applicability to the entire European Economic Area; therefore, also Norway operates under the legal framework of the GDPR.

47 A.K. Befring, 'Norwegian Biobanks: Increased Complexity with GDPR and National Law', in: S. Slokenberga, O. Tzortzatou, and J. Reichel (eds), *GDPR and Biobanking: Individual Rights, Public Interest and Research Regulation across Europe, Law, Governance and Technology Series* 43 (Cham: Springer International Publishing, 2021) 323–344. DOI: 10.1007/978-3-030-49388-2_18.

supervisory authority Garante in Italy.[48] Additionally, further processing of health data for research purposes may be against medical confidentiality laws in some jurisdictions, which is particularly relevant for the research reuse of health data that was initially obtained in a clinical context. Collectively, these considerations clearly highlight that ascertaining the compatibility of further processing within the meaning of Article 5 GDPR is a necessary but not a sufficient condition for enabling further processing. Even where further processing meets the criteria for compatibility, it may still not be legally permissible due to other constraints.

It is noteworthy to recall that, under the GDPR, the presumption of compatibility operates only if and when the safeguards provided for in Article 89(1) GDPR are duly applied. Namely, it must be ensured that "that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation". Where possible, secondary purposes should be achieved "by further processing which does not permit or no longer permits the identification of data subjects", or by "pseudonymisation provided that those purposes can be fulfilled in that manner". When transferring data to a third-party recipient for the recipient's own research purposes, it is unclear whether ascertaining the actual operationality of these safeguards is part of the compatibility test to be performed by the original data controller or whether assurances about them in suitable forms from the recipient would suffice for the compatibility test of the transferring data controller.

6.1 *Lawful Basis for Further Processing of Personal Data for Scientific Research*

As a general rule, the GDPR mandates that any data processing activity carried out using personal data must be based on one of the six conditions listed in Article 6(1) of the Regulation, often referred to as legal bases or lawful bases for data processing. In practice, this means that a data controller deciding on the nature of data processing must choose the most appropriate legal basis prior to commencing the processing.[49] However, the applicability of this requirement to the cases of compatible further processing is not straightforward. According to Recital 50 of the GDPR, when further processing of data is deemed compatible with the primary purpose of data processing, "no legal basis separate from

48 Section 110-a, PERSONAL DATA PROTECTION CODE Containing provisions to adapt the national legislation to 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; Text released on 26 March 2020.

49 *Supra* note 19.

that which allowed the collection of the personal data is required". This also applies to further processing for scientific research purposes, which "should be considered to be compatible lawful processing operations" (Recital 50 GDPR).

An intuitive interpretation of the statement "no legal basis separate from that which allowed the collection of the personal data is required" could be that the legal basis for data collection by the data controller also automatically applies to the further processing that has been deemed compatible. However, it is crucial to emphasise that this sentence only applies to the compatible further processing carried out by the data controller who already holds the data. This controller can indeed continue relying on the existing legal basis to process the data for research purposes. However, any downstream data controller (i.e., a data controller who receives existing data from another controller, as opposed to the data subject) must establish its own legal basis for its primary processing. As such, the collection by a downstream controller effectively marks a "reset" in the chain to the extent that it takes place for a primary purpose, which needs a legal basis on its own. Moreover, even for the initial data controller this interpretation only holds true where certain conditions are met. A closer reading of Recital 50 GDPR provides a valuable insight into the potential constraints to the transferability of the initial legal basis to the further processing. In particular, Recital 50 states that the data controller should undertake a formal compatibility test for further processing "after having met all the requirements for the lawfulness of the original processing". One of the necessary criteria for lawful data processing is to ascertain that the data processing operation fulfils the requirements of the selected legal basis, in accordance with Article 6(1) GDPR. It follows that the existing legal basis used for the primary data processing cannot automatically cover further processing unless the intended further processing operation fulfils the requirements of the legal basis as spelled out in Article 6(1). This conclusion is in line with Article 5 GDPR where legitimacy, Article 5(1)(a), and purpose limitation, Article 5(1)(b), are cumulative requirements. Neither Article 5 nor any other article in the GDPR suggests that the requirement for lawfulness in Article 5(1)(a) is not applicable where the processing complies with Article 5(1)(b), the compatibility of purposes. In particular, Article 6(4) GDPR clearly suggests that prior to performing the compatibility test, the controller first needs to establish the legal basis (other than consent and legal obligation) for the purpose of the further processing. As the recitals of the GDPR are non-binding and are only meant to support the interpretation of the actual legal provisions, any preamble-based conclusion overriding principles such as the requirements of a legal basis under Article 6 GDPR is hard to justify. Also, Article 8(2) of the EU Charter of Fundamental Rights spells out as an absolute requirement that the

processing must take place for a specified purpose and be based on either consent or a legitimate basis laid down by law, thus adding further to the assumption that compatibility of purpose is not sufficient to ascertain legitimacy.

In the context of scientific research, including biomedical research, there is only a subset of options for the legal basis the data controller can choose from: consent 6(1)(a) GDPR; legal obligation 6(1)(c) GDPR; task carried out in the public interest 6(e); legitimate interests 6(1)(f) GDPR.[50] In the case of research related to health and genetic data, national law must foresee an implementation of suitable Article 9(2) GDPR options or the data controller must rely on consent. The availability of choices can be further limited by factors such as national implementation of the GDPR in the data controller's country.[51] For example, the legal basis 6(1)(e), performance of a task carried out in the public interest, may be one of the more suitable legal bases for biomedical research. However, this option is only available to data controllers in those European countries where the national implementation of the GDPR has resulted in a law conferring a clear mandate upon the controller to perform research in the public interest.[52] This uneven availability of the Article 6(1)(e) GDPR legal basis may change in the future, particularly with the expected adoption of the EHDS Regulation. If endorsed by European legislators, the EHDS Regulation could become a reference Union law based on which personal health data may be shared under Article 6(1)(e) GDPR.

The limited choices of valid legal bases that a particular controller can use for scientific research clearly suggests that there will be cases where the original legal basis cannot be extended to cover further processing for scientific research: for example, where data was collected in a healthcare context based on a treatment contract (Article 6(1)(b) GDPR). As a consequence, the controller will need to select a different legal basis and ensure that the requirements associated with this legal basis are fulfilled. Where the controller cannot rely on a national law for its processing (Article 6(1)(c) or 6(1)(e) GDPR), either consent has to be obtained, or a balancing test needs to be performed to justify processing based on legitimate interest (Article 6(1)(f) GDPR). Such a balancing test entails asking questions that are similar to those posed as part of a compatibility test. Thus, where the legitimate interest of the controller

50 See, e.g., EDPB Document on response to the request from the European Commission for clarifications on the consistent application of the GDPR, focusing on health research, adopted on 2 February 2021, para 12.

51 *Ibid.*, para 13.

52 R. Becker, A. Thorogood, J. Ordish and M.J.S. Beauvais, 'COVID-19 Research: Navigating the European General Data Protection Regulation', *Journal of Medical Internet Research* 22(8) (2020) e19799. DOI: 10.2196/19799.

(Article 6(1)(f) GDPR) is the legal basis for scientific research, the controller will be carrying out an assessment that significantly overlaps with the compatibility test under Article 6(4) GDPR.

Besides ensuring that further processing of data for scientific research purposes is grounded in a valid legal basis, the data controller needs to demonstrate compliance with other GDPR requirements. These include compliance with the principles outlined in Article 5 GDPR, adoption of institutional and technical safeguards described in Article 89(1), and, when processing personal health data, additionally fulfilling one of the conditions listed under Article 9(2) of the Regulation. Finally, it is worth mentioning that there may also be sector-specific laws bearing on further research processing of health data that impose additional compliance requirements. For example, medical secrecy or confidentiality laws in a particular country may limit data controllers' ability to undertake biomedical research by further processing personal health data of individuals.

Collectively, these considerations clearly indicate that even though Article 5(1) GDPR and Recital 50 appear to afford special privileges to further processing of data for scientific research, in practice, numerous barriers and compliance requirements exist. It is important that institutions processing personal health data for research purposes are aware of these limitations so that they are not lulled into a false sense of security by the presumption of compatibility of further processing.

7 Discussion

Based on the above analysis of the GDPR, we derive from Article 5(1)(b) GDPR that any instance of data collection by a data controller always marks the beginning of primary processing, irrespective of whether the data are collected directly from the data subject or obtained from another (upstream) controller. According to Article 5(1)(b) GDPR, "Personal data shall be [...] collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes". This is a general requirement that must be respected independently of whether data are newly generated or already existing. Therefore, "initial purpose" in Article 5(1)(b) GDPR, in our opinion, refers to any collection by a controller for one or more specific purposes, whether this is a primary collection (e.g., directly from the data subject) or an ancillary collection (i.e., obtaining existing data from another source). Consequently, any processing for a purpose for which a particular data controller has collected the data does not constitute further processing.

However, we acknowledge that there is currently no consensus over the interpretation of Article 5(1)(b) GDPR, and some members of the European data protection community may hold different views regarding this matter. A notable competing interpretation of further processing is a data-focussed view. According to this view, when distinguishing further processing from primary processing, controllers need to consider the entire data lifecycle, from the generation to the deletion of data.[53] This approach also mirrors the manner in which primary and secondary uses of health data are commonly discussed in the context of biomedical research, which enhances its intuitive appeal. The data-focussed approach sees the first collection (i.e., directly from the data subject) for specific purposes as "initial" and therefore implies that all downstream processing for different purposes, regardless of the identity of the controller, is further processing under the GDPR. This conclusion is supported by an easily misleading phrasing in Recital 50, which refers to "processing of personal data for purposes other than those for which the personal data were initially collected". We believe that the ambiguity of the phrasing "initially collected", introduced by Recital 50, has led in some countries to legislation that treats all additional data processing for opportunistic/ancillary purposes as further processing, even where processing is carried out by a downstream data controller.[54] Statements from the EDPS seem to suggest that the EDPS assumes the data lifecycle-based view as well.[55] The EDPB also seems to support a data lifecycle approach where they give examples for further processing in the context of "primary and secondary usage"[56] but ultimately, clarification is supposed to be provided in the Guidelines on Scientific Research, expected to be published in the near future. The previous Opinion on Purpose Limitation refers to the definition in the Directive 95/46/EC, which differs from the GDPR.[57]

53 K. Pormeister, 'Informed Consent to Sensitive Personal Data Processing for the Performance of Digital Consumer Contracts on the Example of "23andMe"', *Journal of European Consumer and Market Law* 6(1) (2017) 17–23, available online at https://kluwerlawonline.com/journalarticle/Journal+of+European+Consumer+and+Market+Law/6.1/EuCML2017004 (accessed 5 November 2021).

54 E.g., Belgian Data Protection Act Section 2; Italian Data Protection Act, Chapter II, Section 2-b, para 3.

55 The EDPS discusses the compatibility of processing referring to "the original or a new controller". *Supra* note 45, section 6.7.

56 European Data Protection Board (EDPB), *Guidelines 03/2020 on the Processing of Data Concerning Health for the Purpose of Scientific Research in the Context of the COVID-19 Outbreak*, April 2020.

57 Recital 28 of the DPD says "whereas the purposes of processing further to collection shall not be incompatible with the purposes as they were originally specified". This has led to the understanding that further processing could be seen as any processing following the

Another potential source of misunderstanding in Recital 50 is the phrasing that in case of compatibility, "no legal basis separate from that which allowed the collection of the personal data is required." This statement is often read as an unconditional transferability of the legal basis of the collection to compatible further processing. Examples include guidance documents from multiple data protection authorities.[58] We found only one source acknowledging that the requirement to have a valid legal basis for the collection needs to be met for further processing as well in order for the controller to be able to continue relying on that legal basis for further processing.[59] However, several stakeholders have expressed the view that the statement on the legal basis is not reflected in any article of the GDPR and that the requirements of Article 5 GDPR should be understood as cumulative, also referring in this context to Article 8(2) of the Charter.[60] In this context, the sentence that "Further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes should be considered to be compatible lawful processing operations" gives rise to the question of what "lawful" refers to. Kazemi suggests that the sentence may be an editorial mistake and a remnant from the draft version.[61] Indeed, the proposal for the GDPR by the European Commission[62] had a provision under Article 6 that processing necessary for scientific research

collection. (Article 29 Data Protection Working Party, 'Opinion 03/2013 on purpose limitation', Section III.2.1).

58 Example 1: UK Information Commissioner (ICO), Guide to Data Protection/Guide to the General Data Protection Regulation (GDPR)/Principles/Purpose limitation: "If your new purpose is compatible, you don't need a new lawful basis for the further processing", available online at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/principles/purpose-limitation/ (accessed 23 March 2022).
Example 2: Finnish Data Protection Ombudsman: "The controller's processing of personal data for compatible purposes can be based on the same processing basis as the original processing, in which case a new basis is not required", available online at https://tietosuoja.fi/en/defining-the-research-scheme-and-purpose-for-processing-personal-data (accessed on 23 March 2022).

59 H. Lovells, 'Compatibility Test: Can I Process Lawfully Collected Personal Data for a New Purpose?', *JD Supra* (2021), available online at https://www.jdsupra.com/legalnews/compatibility-test-can-i-process-6136680/ (accessed 23 March 2022).

60 See, e.g., *supra* note 45, and Article 29 Data Protection Working Party, 'Opinion 03/2013 on purpose limitation'.

61 R. Kazemi, *General Data Protection Regulation* (*GDPR*), 1st edn. (Hamburg: tredition, 2018).

62 Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 2012/0011 (COD), legislative proposal by the European Commission, available online at http://www.europarl.europa

would always be lawful if in compliance with the conditions and safeguards for scientific research (then Article 83, now Article 89 GDPR), a provision that was still present in the draft of the Council of the European Union in June 2015.[63] However, lawfulness has a much wider scope under the GDPR (e.g., for processing special categories of data or for transfer to third countries), and may also be affected by other legislations such as criminal and civil law provisions in national regulatory regimes. One could also argue that "lawful" in this case merely emphasises the lawfulness with respect to the purpose limitation, but does not replace other legitimacy required under the GDPR.

The ambiguity arising from the phrasing of the GDPR has been observed by the UK Government, which has highlighted the need for clarifying both the question of transferability of the legal basis for compatible processing as well as whether processing by downstream controllers constitutes further processing.[64] We share the view that clarifications are much needed and we hope that with our analysis, we are able to contribute valuable insights to that end.

The discussion of further processing, compatibility, and in particular the privileged status of scientific research often gives rise to concerns over insufficient protection of data subjects in relation to the processing of their data. The EDPS has expressed concern that the compatibility by default of scientific research may lead to a carte blanche that could be abused, resulting in an insufficient protection of the rights and freedoms of data subjects.[65] However, these concerns may not be wholly justified, given that, as we have demonstrated, additional safeguards are in place. These stem from the requirements that the initial controller pursuing further processing needs to establish a valid legal basis under Article 6(1) GDPR and demonstrate compliance with Article 89(1) GDPR in order to ascertain compatibility of further processing. If controllers can rely on a national or EU law giving them an explicit mandate to carry out research in the public interest or if they have a legal obligation to do so, this law

.eu/RegData/docs_autres_institutions/commission_europeenne/com/2012/0011/COM_COM%282012%290011_EN.pdf (accessed 23 March 2022).

63 Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 2012/0011 (COD), general approach adopted in the Council, 11 June 2015, available online at https://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf (accessed 23 March 2022).

64 Data: a new direction; Public consultation on reforms to the UK's data protection regime (2021), available online at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1022315/Data_Reform_Consultation_Document__Accessible_.pdf (accessed 23 March 2022).

65 *Supra* note 42.

demonstrates, *inter alia*, a societal interest in the processing. Such weighing of interests is in line with the spirit of the GDPR, which emphasises the balancing of personal and public interests to a much greater extent than its preceding Directive.[66] Where controllers cannot rely on a law for the further processing, they either need to have a valid consent from data subjects or perform a balancing test to justify a processing under Article 6(1)(f) GDPR, with the latter being comparable to the compatibility test. However, Article 6(1)(f) GDPR cannot be relied upon by public authorities where data processing is related to the performance of their tasks. Article 89(1) further emphasises the requirement that the implementation of safeguards for data minimisation have to be demonstrated for scientific research exemptions. In particular, where data are shared for a third party's research, that third party has to provide sufficient guarantees to the disclosing party that its research processing will actually fall under scientific research and will also comply with the applicable Article 89(1) GDPR requirements. In the case of processing special categories of data, implementations of Article 9(2)(j) GDPR also have to be considered for legitimate processing. We therefore conclude that the risk of excessive processing for research, based on the privilege in Article 5(1)(b) GDPR, is limited.

8 Conclusion

In this article, we have analysed secondary use of personal health data through the lens of the GDPR. We have focused on the related GDPR concept of further processing and discussed its implications for the compliance obligations of healthcare and research institutions. When elucidating the similarities and differences between secondary data use and further processing, we conclude that the latter is to be understood in relation to the purpose for which a particular data controller has collected the data. Under this "data controller"-centred view, further processing can only be performed by a data controller that already has the data, as opposed to a new controller seeking access to the data. When data are transferred between two controllers, both controllers are required to establish a valid legal basis under Article 6(1) GDPR and where applicable under Article 9(2) GDPR, ensuring that the legitimation is appropriate for the controller's role and interest in the processing. The selection

66 GDPR Recital (4): "The processing of personal data should be designed to serve mankind. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality."

of these legal bases must be informed by contextual factors relevant to the intended further processing. As such, the most appropriate legal basis/bases may or may not be the same as the legal basis currently relied upon by the data-transferring (or upstream) controller.

We also discussed the special case of further processing for scientific research purposes. Several provisions in the GDPR lend themselves to the interpretation that further processing of personal data for research purposes is compatible with the primary purpose of data processing, that is, the purpose for which the controller originally collected the data. However, it must be emphasised that compatibility of further processing is only one component of, and should not be equated with, the lawfulness of further processing. To ensure lawfulness of processing, controllers seeking to further process data for scientific research purposes will need to meet additional requirements, such as identification of a suitable legal basis as per Articles 6(1) and 9(2) GDPR, and compliance with the principles of the regulation. Furthermore, controllers must be able to demonstrate compliance with any other obligations defined in an applicable Union or Member State law, as per Article 89(1) GDPR or Article 9(4) GDPR. While these may pose practical barriers to the intended further processing, together they provide a robust set of safeguards for the protection of the rights and freedoms of data subjects.

Acknowledgements

The work of RB was supported by ELIXIR-Luxembourg. The work of DC was supported by the European Union's Horizon 2020 research and innovation programme Coordination and Support Action HealthyCloud (grant agreement no. 965345) and the Innovative Medicines Initiative 2 Joint Undertaking Research and Innovation Action European Platform for Neurodegenerative Diseases (EPDN, grant agreement no. 101034344). The work of GC was supported by the Italian Health Ministry under the activity "Strategia Genomica italiana: istituzione di una cabina di regia a supporto dell'iniziativa europea 1+Million Genomes (1+MG) e Beyond 1+MG (B1MG) e del Coordinamento Interistituzionale per la Genomica in Sanità Pubblica." The work of FMG was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — NFDI 1/1 "GHGA — German Human Genome-Phenome Archive". The work of PN was supported by the Instituto de Salud Carlos III, Spain under "Infraestructura de Medicina de Precisión asociada a la Ciencia y Tecnología (IMPaCT) de la Acción Estratégica en Salud 2017–2020. IMP/00009." and by "Research Group of the Basque University System: Social and Legal

Sciences applied to the New Technosciences (IT1541)". The work of AT was supported by the European Union's Horizon 2020 research and innovation programme Coordination and Support Action Beyond 1 Million Genome (B1MG, grant agreement no. 951724). FMG is a member of the European Group on Ethics in Science and New Technologies. This article is written in a purely private capacity and the views expressed here cannot be attributed to anyone other than the authors.