European Journal of Human Genetics (2018) 26:149–156
https://doi.org/10.1038/s41431-017-0045-7

REVIEW ARTICLE

Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation

Mahsa Shabani[1] · Pascal Borry[1]

Received: 17 July 2017 / Revised: 26 September 2017 / Accepted: 31 October 2017 / Published online: 29 November 2017

Abstract

Genetic data contain sensitive health and non-health-related information about the individuals and their family members. Therefore, adopting adequate privacy safeguards is paramount when processing genetic data for research or clinical purposes. One of the major legal instruments for personal data protection in the EU is the new General Data Protection Regulation (GDPR), which has entered into force in May 2016 and repealed the Directive 95/46/EC, with an ultimate goal of enhancing effectiveness and harmonization of personal data protection in the EU. This paper explores the major provisions of the new Regulation with regard to processing genetic data, and assesses the influence of such provisions on reinforcing the legal safeguards when sharing genetic data for research purposes. The new Regulation attempts to elucidate the scope of personal data, by recognizing pseudonymized data as personal (identifiable) data, and including genetic data in the catalog of special categories of data (sensitive data). Moreover, a set of new rules is laid out in the Regulation for processing personal data under the scientific research exemption. For instance, further use of genetic data for scientific research purposes, without obtaining additional consent will be allowed, if the specific conditions is met. The new Regulation has already fueled concerns among various stakeholders, owing to the challenges that may emerge when implementing the Regulation across the countries. Notably, the provided definition for pseudonymized data has been criticized because it leaves too much room for interpretations, and it might undermine the harmonization of the data protection across the countries.

Background

Recent advancements in genomics and bioinformatics have led to vast amounts of genomic data being generated in clinical and research settings. In order to obtain a better understanding of these data and identify potential correlations between diseases and underlying genetic factors, sharing genomic data in research and clinical settings is deemed necessary [1, 2]. In the view of increasing data-sharing practices, the importance of adopting adequate legal protection for data subjects when using individual-level genomic data has been stressed. Sharing identifiable genomic data is a form of processing personal data, and as such would fall within the scope of data protection laws [3].

Genetic data contain unique information about the data subjects and their blood relatives, highlighting the significance of adopting adequate privacy protection measures when processing genomic data [4, 5]. Adopting adequate privacy protections for genomic data has been endorsed by the establishment of the International Declaration on Human Genetic Data, which was issued on 16 October 2003 by UNESCO as complementary to its Universal Declaration on Human Genome and Human Rights from 11 November 1997.

In the light of international human rights, regimes that endorse privacy rights in general and genomic privacy rights in particular, laws, and regulations at the EU level have been established in order to provide enforceable legal instruments in protecting the privacy of individuals. In the European Union, protection of personal data has been pursued by establishing the Directive 95/46/EC 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 (hereafter the "Directive"). The Directive was established in order to ensure the lawful and fair processing of personal data via information technology. The Directive is only meant to

✉ Mahsa Shabani
mahsa.shabani@kuleuven.be

[1] Centre for Biomedical Ethics and Law, Department of Public Health and Primary Care, University of Leuven, Kapucijnenvoer 35 blok d—box 7001, 3000 Leuven, Belgium

apply to "personal data" and is meant to exclude data that are not "directly or indirectly" identifiable or that are considered to be anonymous. The Directive stipulates that the processing of personal data should not be incompatible with the original purposes of data collection and that the data should only be kept for as long as is necessary to achieve those purposes.

In 2009 the European Commission embarked on mission to reform the Directive. The ultimate goal of the reform was to make the Directive more effective with regard to the advancements in information communication technologies, which have remarkably transformed collection, storage, and transfer of high volumes of data across borders. In addition, the Directive could not introduce harmony and consistency in the data protection realm in the EU, as it was transformed into national laws and this resulted in 27 different, national, versions of the Directive. Therefore, the replacement of the Directive by a new Regulation, which is directly enforceable in all member states has been pursued. In January 2012 the European Commission released a "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". There were later amendments to the Proposal voted on by the European Parliament on 12 March 2014. After this, the Council agreed to a common approach on a revised text of the Proposal on 15 June 2015 and a period of trialogue between the three EU bodies (Commission, Parliament, and Council) started. After negotiations between the three EU bodies, on 15 December 2015 the European Parliament, the Council, and the Commission reached an agreement on the new data protection rules. The EU General Data Protection Regulation (hereafter the "Regulation") has been introduced with the ultimate goals of harmonizing data protection across the EU, and facilitating the flow of information across borders and enhancing privacy protection. On 4 May 2016, the official text of the Regulation was published in the EU Official Journal in all the official languages. While the Regulation entered into force on 24 May 2016, it shall apply from 25 May 2018.

In this paper, we will analyze impact of four elements within GDPR on the processing of genomic data for research purposes. These elements include the definition and scope of personal data; recognition of genetic data within the special categories of personal data; processing personal data under the research exemption; and conditions and safeguards for processing data under research exemptions. To this purpose, we will critically review the pertinent provisions on the GDPR in contrast to the relevant provisions of the Directive. Our discussions will benefit from the arguments provided by the exiting commentaries, and position statements of research organizations and professional bodies.

Definition and scope of personal data

The concept of "personal data" is a key concept in the framework of the Regulation. Once data have been recognized as personal data under the Regulation, processing of the data should be pursuant to the main principles laid out in Article 6. Previously, the definition provided by the Directive has been criticized because of a lack of clarifications regarding the scope of personal data in a number of aspects, including a distinction between anonymized vs. anonymous data [6] and the status of key-coded or pseudonymized data.

In the definition provided by GDPR, the core elements of the definition from the Directive have been maintained, mainly defining personal data as "any information relating to an identified or identifiable natural person ("data subject")". However, in the catalog of identifiers, the definition provided by the Regulation includes "genetic" (Article 4.1), which was not included in the Directive's definition of personal data. Although "genetic" has been generally included as an example of identifier factors, one can consider that this will only apply to identifying genetic factors [7].

Furthermore, the Regulation does not distinguish between *anonymized* and *anonymous* data, when explaining the scope of the personal data in Recital 26. Previously, a distinction between anonymous data (data that never were identifiable) and *anonymized* data (data that were rendered anonymous) has been proposed in the literature. Beyleveld argues that rendering personal data anonymous should indeed be considered as "processing" data. Therefore, such data should fall within the scope of data protection regulation and the act of anonymization should be considered "processing" for the purpose of data protection regulations [8]. This approach resonates with the advice from the Article 29 Data Protection Working Party (hereafter the "Working Party"; the Article 29 Data Protection Working Party, which has been set up under this Directive, is a group that regularly issues statements on matters relevant to the Directive and has been highly influential in providing interpretations for the Directive's provisions. The Working Party is composed of representatives from the Member States' Data Protection Authorities, the EU Commission, and the EU Data Protection Supervisor, which is an independent authority), which states "Anonymization constitutes a further processing of personal data; as such, it must satisfy the requirement of compatibility by having regard to the legal grounds and circumstances of the further processing". The GDPR therefore excludes processing data for statistical or research purposes from the scope of data protection, if the data are rendered anonymized. Important implication of this provision will be that individuals will not be entitled to data protection rights, if their data are collected in identifiable manner but later rendered anonymized. One example is when data are collected in a clinical setting in an

identifiable manner, and anonymized later on to be used for various purposes either by private or public parties. Although anonymized data are considered non-personal for the purpose of GDPR, still individuals may be concerned that how the data extracted from them will be used and for which purposes.

Pseudonymization

For the first time, the Regulation defined the concept of pseudonymization. According to Article 4(5), "*Pseudonymization" means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to an identified or identifiable natural person*". Considering the existing controversies around the status of pseudonymized data for the purpose of data protection regulation, and the diversity in approaches toward pseudonymization [9], the efforts made in the new Regulation to delineate the concept are particularly important for data-sharing practices important.

In Recital 26, the Regulation asserts that pseudonymized data should be considered personal data if it could be attributed to a natural person by the use of additional information. Moreover, in the assessment of the identifiability of the data "all the means reasonably likely to be used, such as singling out, either by the controller or by another person" should be taken into consideration. This will open the door for varying interpretations on what would constitute the "all the means reasonably likely to be used", and how the criteria for identifiability should be determined. It is conceivable that the pseudonymization of data, if accompanied by appropriate measures that make re-identification unlikely, renders data anonymized or result in adopting lighter regulatory provisions in comparison to identifiable data [10]. This approach resonates with the Article 29 Working Party opinion on the concept of personal data: "… using a pseudonym means that it is possible to backtrack to the individual, so that the individual's identity can be discovered, but then only under predefined circumstances. In that case, although data protection rules apply, the risks at stake for the individuals with regard to the processing of such indirectly identifiable information will most often be low, so that the application of these rules will justifiably be more flexible than if information on directly identifiable individuals were processed" [11].

Recognizing pseudonymized data in the Regulation as personal data will affect the practices of those research studies that are currently considering pseudonymized data as non-personal data. One example is epidemiological research, which extensively use key-coded or pseudonymized data, and, depending on the applicable national laws, currently considers pseudonymized data as non-identifiable. As Van Veen points out: "As pseudonymized or key-coded data are the working vessel of registry-based research, this new definition of personal data could have very negative consequences for research. It would mean that one would have to fall back on the research exception in many more cases than at present, with all the bureaucracy that might be attached to the permission for use of the exception" [12]. This is expected to be a significant change in Member States such as the Netherlands, where pseudonymized data, under certain conditions, have been considered to fall outside the scope of the definition of personal data [13]. Similarly, as Rumbold and Pierscionek point out, "The United Kingdom Information Commissioner's Office currently treats pseudonymized data as anonymous where it is used by a third party who does not possess the requisite key code." [14] Indeed, the lack of clear provisions in the Directive toward pseudonymized data allowed for broad interpretations in Member States' laws of the scope of such definitions. In addition, the existing heterogeneity in pseudonymization methods used across Member States could be seen as a potential challenge in implementing the pertinent provisions concerning pseudonymization and hinder cross-border genomic data-sharing [13].

Recognition of genetic data within the special categories of personal data

Regulation has marked certain categories of personal data as sensitive, and this entails higher protection and stricter requirements for the processing of such data. According to Recital 51: "*Personal data which are, by their nature, particularly sensitive in relation to fundamental rights and freedoms merit specific protection as the context of their processing could create significant risks to the fundamental rights and freedoms.*"

Recognizing special categories of data by the Regulation was not unprecedented, as the Directive has adopted a similar approach on this matter. Article 8 of the Directive contained a general prohibition on processing personal data that reveal racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, and the processing of data concerning health or sex life. As is further explained by Working Party in the Advice paper on Special Categories of Data (sensitive data), the definition included not only data that by its nature contains sensitive information, but also data from which sensitive information with regard to an individual could be concluded.

GDPR, in contrast, explicitly recognizes the sensitive nature of genetic data collected in a variety of settings.

In Article 9, an adjusted definition of special categories of personal data has been provided that includes genetic data and biometric data, among others. Inclusion of genetic data in the catalog of sensitive data is in line with the opinion of majority of Working Party members. At the national level, some Member States included genetic data and biometric data in their catalog of special categories of personal data [15]. Establishing stricter requirements for processing genetic data seems appropriate, in the view of the heightened concerns regarding potential misuses of genetic data, which could result from increased availability of genetic data.

In addition, Article 4(13) provides a definition for genetic data and in Recital 34 further explains that, "analysis of biological sample" includes in particular chromosomal, deoxyribonucleic acid (DNA), or ribonucleic acid (RNA) analysis, or the analysis of other elements that enables equivalent information to be obtained. This definition implies that not only genetic information that drive from DNA materials, but also genetic information that could result from analysis of other materials such as molecular and biological materials will be recognized as genetic data for the purpose of GDPR. The questions remain about the status of other types of genetic information that may not result from analysis of biological materials, but other sources such as "genealogical information gathered through various questionnaires" [16].

A point to consider is how to distinguish genetic data from the biological material from which they are derived. Such clarification is particularly important for biobanks and those researchers who aim for sharing biological samples that potentially contain genetic information. The Regulation and Articles delineating the scope of the Regulation do not discuss this point. In the absence of clear provisions in the Regulation concerning biological samples, one way to achieve clarity is to look to interpretations. One approach is since the ultimate intention of the Regulation is to protect personal data, a broad interpretation should be applied, which could allow for the inclusion of all sources, including biological samples that contain genetic data. However, given the definition provided for genetic data in the Regulation which explicitly states "data" (not samples), it will be hard to maintain such a broad interpretation [17].

Processing personal data and special categories of data under the research exemption

Processing sensitive data under specific conditions has been addressed in Article 8 of the Directive. Accordingly, sensitive personal data could be processed if the explicit consent of the data subject has been obtained. Otherwise, the processing of sensitive data could be carried out "for reasons of substantial public interest", if "suitable safeguards" were in place. Although research has not been explicitly included as a reason for processing sensitive data in Article 8, recital 34 of the Directive mentions scientific research as a potential example of "reasons of substantial public interest" that could be utilized by the Member States when implementing the Directive.

In practice, processing sensitive data under the exception of "public interest" has been done under strict conditions, which were set by the Member States. However, fulfilling such conditions appeared burdensome, thus rendering the processing under "public interest" exception less favorable. As Paul Quinn notes: "Whilst the public interest option in Directive 95/46/EC allows states to legislate for the possibility of using personal health data for scientific research without consent, the conditionality that is required means that such options cannot be considered as "constraint-free". Imagine for instance conditionality that requires an extremely high level of pseudonymization. Another requirement may (depending upon the jurisdiction in question) require that approval is sought and obtained from a national, regional, or organizational ethics body" [18].

Adopting a new approach toward processing personal data for research purposes was one of the most controversial topics in the course of making the new Regulation. While the Commission's proposal was similar to the Directive's approach for processing personal data, the later amendments voted on by the European Parliament on 12 March 2014 laid out considerably stricter conditions for such processing. According to the amended version of the Parliament, in the absence of consent from data subjects, processing of data concerning health for research purposes should only be allowed if it serves a "high public interest" and if "that research cannot possibly be carried out otherwise" (Article 81(2a)). The proposed amendments on the Commission's draft by the Parliament fueled massive concerns among the biomedical and health research community, who saw the proposed requirements as a barrier to research [19–21]. The pertinent provisions, and especially Articles 81 and 83 concerning the use of health data including genetic data for research purposes, were extensively discussed by the European Council afterwards. Ultimately, the final text of the Regulation adopted a more research-friendly approach. According to the new Regulation, a "research exemption" has been recognized under a number of Articles.

First, while processing special categories of data has been generally prohibited, Article 9.2(j) of the Regulation permits processing of special categories of personal data when it is necessary for *archiving purposes in the public interest*, *scientific or historical research purposes, or statistical purposes* in accordance with article 89(1).

This could occur without the explicit consent of the data subject having been obtained as long as this is permitted under EU or Member State law and appropriate safeguards are in place. It should be noted that the GDPR recognizes the challenges of obtaining specific consent for research purposes at the time of data collection, therefore provided that data subjects should be allowed to give consent to certain areas of scientific research (Recital 33). The Regulation states that "Member States may maintain or introduce further conditions, including limitations, with regard to the processing of genetic data, biometric data or data concerning health" (Article 9.4). Member States, therefore, could aim for stronger protections for genetic data by requiring stricter conditions for processing genetic data for research purposes. However, maintaining varying requirements by Member States will undermine a goal of harmonization of legal framework for processing genetic data within the EU. This is particularly challenging, given the importance of collaborative genetic research, which entails cross-border processing of genetic data. On a similar note, in Recital 53, the Regulation warns Member States that they should not use this discretion in a way that "hamper the free flow of personal data within the Union when those conditions apply to cross-border processing of such data."

Second, the research exemption could provide a legal basis for the secondary processing of personal data, something that could also be provided by the "further processing" provisions. Accordingly, Recital 50 indicates that further processing for archiving purposes in the public interest, for scientific and historical research purposes or for statistical purposes should be considered to be compatible processing. This means that retrospective use of genetic databases will be allowed, thus optimizing the use of already collected data for future research purposes. However, where further processing of personal data is desired, the principles of transparency and fairness should be respected. In particular, the data subjects should receive the relevant information regarding that further processing in advance (Article 13(3)). Therefore, researchers who obtained the data from the data subjects and aim for further processing of data for scientific research purposes should ensure that the data subjects receive the relevant information prior to further processing. Such requirement shall not apply "where the provision of information to the data subject proves to be impossible or would involve a disproportionate effort" (Recital 62).

Third, Article 6 lays out the grounds for the lawful processing of personal data without consent, including but not limited to the condition when "processing is necessary for the purposes of the legitimate interest". A similar approach has been adopted in the conditions set for transferring personal data to third countries under Article 49(1), where that transfer can be carried out in the absence of consent when "necessary for the purposes of compelling legitimate interests pursued by the controller that are not overridden by the interests or rights and freedoms of the data subject."

Although processing for research purposes is not explicitly listed under "legitimate interests", the further explanation provided by Recital 47 and Recital 113 could potentially provide sufficient grounds to process personal data for research purposes. This interpretation resonates with the opinion of the Article 29 Working Party on the notion of Legitimate Interest, where this opinion includes scientific research as a legitimate interest (subject to appropriate safeguards) [22].

Concerning the definition of scientific research, it is worth noting that the Regulation favors a broad interpretation, which will effectively broaden the scope of national laws [23]. According to Recital 159, "the processing of personal data for scientific research purposes should be interpreted in a broad manner including, for example, technological development and demonstration, fundamental research, applied research, and privately funded research." Therefore, both private and publicly funded research could benefit from the research exemption provisions under the Regulation. However, concerns regarding potential misuse of research exemption by commercial actors has led some such as Biobanking and BioMolecular resources Research Infrastructure—European Research Infrastructure Consortium (BBMRI-ERIC)'s to argue in favor of specifying research exemption to scientific research seeking public interest [24]. Since the contribution of both public and private entities in advancement of biomedical research is essential, the GDPR's approach in inclusion of both private and publicly funded research may be seen beneficial as far as the research's objectives align with the public interests and the procedure of data processing is transparent.

Conditions and safeguards for processing data under the research exemption

Derogations from data subjects' rights have been introduced when processing data for scientific research purposes. Article 89(2) of the new Regulation allows Member States to provide for derogations from the rights referred to in Articles 15 (right of access), 16 (right to rectification), 18 (right to restriction of processing), and 21 (right to object). Recital 156, however, provides a longer list of derogations that could be made by Member States, including "*derogations with regard to the information requirements and rights to rectification, to erasure, to be forgotten, to restriction of processing, to data portability, and to object when processing personal data for archiving purposes in the public interest, scientific, or historical research purposes or statistical purposes.*"

Article 89(1) outlines some conditions for processing personal data under the research exemptions. Accordingly, processing of personal data for scientific research purposes, among others, shall be subject to appropriate safeguards. However, it has been primarily left to the Member States to define the term "safeguards". Similarly, in the framework of the Directive the term "safeguards" was mentioned on several occasions; however, a clear definition of the nature of such safeguards was not provided. In response, the Working Party stressed the importance of further delineating the definition of safeguards, and illustrating it with examples: "Organizational and technical safeguards could, for example, include measures such as the introduction of Information Security Managements Systems (e.g., ISO/IEC standards) based on the analysis of information resources and underlying threats, measures for cryptographic protection during storage and transfer of sensitive data, requirements for authentication and authorization, physical and logical access to data, access logging and others."

Article 89(1) mentions "pseudonymization" as a measure that can be taken in order to ensure respect for the principle of data minimization. Given that the use of identifiable data is important at times such as for epidemiological research [21, 25], pseudonymization can be a restricting factor for use of genetic data under research exemption. Importantly, when there are other safeguards in place, then it should be possible to use identifiable data for research purposes, without consent. These safeguards could include governance mechanisms such as obtaining approval from ethics committees or data access committees, who are tasked with making an assessment of research proposals. Such oversight shall take into account considerations of the potential risks for data subjects prior to researchers being granted access to their data, and to ensure the benefits of the research outweigh the associated risks.

Regardless, in order to reduce the risks of re-identification, particularly in processing genomic data, adopting controlled-access models has been widely suggested. Controlled-access or managed-access models would allow maintaining a level of control on downstream uses of the research databases, through conducting access review by specialized access committees that oversee the incoming data access requests and assess them for the purpose of approval or disapproval [26]. Furthermore, the access committees could vet the data users and only grant access to bona fide researchers. As Ohm puts it: "Researchers should be allowed to release full, unscrubbed databases to verifiably trusted third parties, subject to new controls on use, and new penalties for abuse" [27].

Other alternative models include archiving and processing data in safe havens, encryption and key management, and technical and organizational security measures. It is worth noting that the dynamic nature of the field and advancements in bioinformatics call for regular updates to ensure adequate safeguards [28].

Concluding remarks

The Regulation took a similar approach regarding the scope of personal data. However, for the first time, GDPR elucidated the term "pseudonymization" and provided a definition. The Regulation asserts that pseudonymized data are considered identifiable and will fall within the scope of the personal data. Moreover, pseudonymization has been introduced as an example of measures that could be used by data processors when processing sensitive data, such as genetic data, on the basis of the research exemption provision. Although the clarification about pseudonymization is important, some uncertainties still remain regarding the impact of the pertinent provisions on current practices, for instance, in relation to adequate minimum standards of pseudonymization. Moreover, it remains to be seen how this would change governance mechanisms concerning sharing de-identified genomic data, such as the consent and the oversight mechanisms.

The new Regulation has recognized the research exemption for processing personal data on a number of occasions, and therefore presented a research-friendly approach. Accordingly, in the absence of consent, personal data, including sensitive data, could be further processed for scientific research purposes and under the conditions set out in Article 89. Considering the increasing attention being directed toward data-sharing for scientific research purposes, the rules set forth by the Regulation regarding the research exemption are of paramount importance.

The provisions set forth for processing personal data under the research exemption could supplement the existing binary approach toward data protection, namely consent or anonymization. [29] Notably, such a binary approach does not respond to the demands of biomedical research, which needs high volumes of data in a fast and easily accessible manner. According to the Regulation, using data for research purposes and sharing it for downstream uses require adopting further organizational safeguards, which go beyond the consent or anonymization approaches. Nevertheless, the Regulation does not elaborate further on such safeguards, leaving it primarily to the Member States to adopt adequate safeguards and conditions for processing data under the research exemption.

In light of the identified ethical and legal concerns associated with using genetic data for research purposes, we stress the importance of safeguards, which could provide a level of control on further processing of data for research purposes in an on-going manner. This will establish additional controls that limit data access to authorized users.

A similar approach has been adopted in a recent report on the *Collection, linking and use of data in biomedical research and health care* by Nuffield Council on Bioethics, which notes, "Because of the risk of misuse and consequential privacy infringement, de-identification, and consent measures may be supplemented by further governance arrangements" [30]. Competent oversight bodies such as ethics committees and data access committees are in the best place to hold control over the access and use of data. By establishing adequate oversight mechanisms from the outset in the process of personal data processing, the ultimate goal of the new Regulation in terms of "privacy by design" will be facilitated, in which data protection safeguards will be built into the products and services from the earliest stage of development.

However, it is important to ensure the existing and emerging oversight bodies are equipped with adequate expertise regarding using and sharing genomic data and are aware of the associated informational risks. In order to achieve this, soliciting the attitudes of the involved parties regarding the associated risks would be necessary. Thereby, the overall governance of personal data processing will go beyond legal requirements, and will take into account the pertinent individual or social concerns that may not be explicitly outlined in the legal provisions. In addition, the oversight of personal data processing should keep pace with recent developments in the field of data science, bioinformatics, and genetics, among others. The risks associated with emerging technologies and the safeguards in protecting the privacy of data subjects should be treated as moving targets. Otherwise, the safeguards will become obsolete and unable to safeguard data subjects in an adequate manner.

Finally, increasing cross-border data-sharing underlines the importance of the harmonization of legal frameworks concerning personal data protection. One of the main goals of the Regulation has been to achieve this by harmonizing the personal data protection landscape across EU. However, concerns remain regarding the real impact of the Regulation on unifying the individual, national regulations toward processing genetic data for research purposes, across Member States. Arguably, the Regulation still leaves room for varying interpretations, for instance, concerning the safeguards that should be established and also in setting further conditions for processing genetic data on the basis of the research exemption provisions. In a position statement, BBMRI-ERIC stressed the significance of ensuring that "Member State-specific derogations are not invoked to block, delay, or otherwise unduly frustrate cross-border data exchange for research purposes". In addition, negotiating sector-specific codes of conducts by professional bodies is suggested as a way to reach harmonization across EU [24]. Further research could explore how Member States will adjust their national laws in the coming 2 years in preparation for enforcing the Regulation in 2018.

Main points

- Recognizing pseudonymized data as personal data by GDPR introduces clarifications to the status of pseudonymized data. Still, the provided definition leaves room for further interpretations on what are the sufficient methods of pseudonymization and when data are fully considered non-identifiable.
- Allowing Member States' to set further limitations on processing genetic data for research purposes may hamper cross-border processing of genetic data and undermine harmonization of data protection within the EU, if those limitations and conditions vary.
- GDPR emphasized pseudonymization as a safeguard when processing data under research exemption. Other safeguards, such as organizational measures and oversight by competent bodies, should be further utilized as they may better suit to the purpose of governance of research at times.

Acknowledgments This work is kindly supported by the Interfaculty Council for Development Co-operation (IRO) of the University of Leuven and Research Foundation Flanders (FWO).

Compliance with Ethical Standards

Competing interests The authors declare that they have no competing financial interests.

References

1. Knoppers BM. Framework for responsible sharing of genomic and health-related data. HUGO J. 2014;8:1.
2. Hayden EC. Geneticists push for global data-sharing: international organization aims to promote exchange and linking of DNA sequences and clinical information. Nature. 2013;498:16–8.
3. Knoppers BM, Harris JR, Tassé AM, et al. Towards a data sharing Code of Conduct for international genomic research. Genome Med. 2011;3:1.
4. European Commission. Ethical, Legal and Social Aspects of Genetic Testing: Research, Development and Clinical Applications, 2004. Available online at: https://publications.europa.eu/en/publication-detail/-/publication/2159a3bb-92af-4c6e-ab27-7bb0189773f9
5. de Paor A. Regulating genetic information—exploring the options in legal theory. Eur J Health Law. 2014;21:425–53.
6. Beyleveld D, Townend D. When is personal data rendered anonymous? Interpreting recital 26 of Directive 95/46/EC. Med Law Int. 2004;6:73–86.
7. PHG Foundation. Call for Evidence on the Proposed Data Protection Regulation and Directive: Response from PHG Foundation. Available online at: http://www.phgfoundation.org/documents/291_1331894630.pdf

8. Beyleveld D. Privacy, confidentiality and data protection. The SAGE handbook of health care ethics. Edited by Ruth Chadwick, Henk Ten Have and Eric Meslin 2011:95-105. DOI: http://dx.doi.org/10.4135/9781446200971.n10
9. Phillips M, Knoppers BM. The discombobulation of de-identification. Nat Biotechnol 2016;34:1102–3.
10. Wellcome Trust. Analysis: Research and the General Data Protection Regulation - 2012/0011(COD), 2016. Available online at: https://wellcome.ac.uk/sites/default/files/new-data-protection-regulation-key-clauses-wellcome-jul16.pdf
11. Article 29 Working Party. Opinion 4/2007 on the concept of PersonalData, 2007.
12. van Veen E-B. Europe and tissue research: a regulatory patchwork. Diag Histopathol. 2013;19:331–6.
13. Moraia LB, Kaye J, Tasse AM, et al. A comparative analysis of the requirements for the use of data in biobanks based in Finland, Germany, the Netherlands, Norway and the United Kingdom. Med Law Int. 2015;14(4) 0968533215571956.
14. Rumbold JMM, Pierscionek B. The Effect of the General Data Protection Regulation on Medical Research. J Med Internet Res 2017;19:e47.
15. Article 29 Working Party. Advice paper on special categories of data ("sensitive data"), 2011. Available online at: http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2011/2011_04_20_letter_artwp_mme_le_bail_directive_9546ec_annex1_en.pdf
16. Chassang G. The impact of the EU general data protection regulation on scientific research. Ecancermedicalscience 2017;11:709.
17. Hallinan D, De Hert P. Brent Mittelstadt, and Luciano FloridiSpringer International Publishing (Cham, Switzerland, 978-3-319-33523-0, 480 pp.) Many have it wrong–samples do contain personal data: the data protection regulation as a superior framework to protect donor interests in biobanking and genomic research. The Ethics of Biomedical Big Data. Springer; 2016. p. 119–37.
18. Quinn P. The anonymisation of research data—a pyric victory for privacy that should not be pushed too hard by the eu data protection framework? Eur J Health Law. 2016;24:1–21.
19. Dove ES, Townend D, Knoppers BM. Data protection and consent to biomedical research: a step forward? Lancet. 2014;384:855.
20. McCall B. European Parliament supports data protection reforms. Lancet. 2014;383:1115.
21. Di Iorio C, Carinci F, Oderkirk J. Health research and systems' governance are at risk: should the right to data protection override health? J Med Ethics. 2014;40:488–92.
22. Article 29 Working Party. Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, 2014. Available online at: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf
23. Pormeister K. Genetic data and the research exemption: is the GDPR going too far? International Data Privacy Law 2017: 7(2): 137–146.
24. BBMRI-ERIC. Position Paper on General Data Protection Regulation 2015. Available online at: http://www.bbmri-eric.eu/wp-content/uploads/BBMRI-ERIC-Position-Paper-General-Data-Protection-Regulation-October-2015_rev1_title.pdf
25. Mascalzoni D, Dove ES, Rubinstein Y, et al. International Charter of principles for sharing bio-specimens and data. Eur J Hum Genet. 2015;23:721–8.
26. Shabani M, Knoppers BM, Borry P. From the principles of genomic data sharing to the practices of data access committees. EMBO Mol Med. 2015;7:507–9.
27. Ohm P. Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Rev. 2010;57:1701.
28. Williams G, Pigeot I. Consent and confidentiality in the light of recent demands for data sharing. Biom J. 2016;59:240–50.
29. Mostert M, Bredenoord AL, Biesaart MC, et al. Big Data in medical research and EU data protection law: challenges to the consent or anonymise approach. Eur J Hum Genet. 2016;24:956–60.
30. Nuffield Council on Bioethics. The Linking and Use of Biological and Health Data; 2013. Available online at: http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf