2498-5473 / USD 20.00
© 2018 Akadémiai Kiadó, Budapest

Hungarian Journal of Legal Studies 59, No 4, pp. 403–419 (2018)
DOI: 10.1556/2052.2018.59.4.5

Big Data and Scientific Research: The Secondary Use of Personal Data under the Research Exemption in the GDPR

János Mészáros*
Chih-hsing Ho**

Abstract. In the age of big data and AI, the ability to extract knowledge and value from personal data is promising, especially for researchers and policymakers. The new findings based on the vast amount of data have the potential to save lives and reduce expenses for the whole society. However, processing sensitive data for a new purpose poses complex ethical, legal and technical challenges. The EU General Data Protection Regulation (GDPR) accounts for this challenge by allowing researchers to process and further use personal data under the 'research exemption'. However, many aspects of this exemption would need further clarification: what level of public interest is necessary e.g., general, important or substantial, how the data should be de-identified and what kind of activities can fit in the definition of 'scientific research'. The issue is elaborated through the GDPR and its implementation in England and Germany, focusing on the secondary use of health data.
Keywords: Big Data, GDPR, secondary use, scientific research, research exemption, privacy, data protection, health data

1. INTRODUCTION

In the past, personal data was mostly collected and processed for a specific purpose. The data protection laws around the world encompassed the purpose and storage limitation principles, which means that the data cannot be used for a new, different purpose and the retention time should not be longer than it is necessary for the specific purpose. In the age of big data and AI, these principles are constantly challenged by researchers and data analytics. Technological advancements and the new algorithms make the purpose limitation principle questionable,[1] especially in the field of scientific research. For instance, the electronic health records became a valuable asset for the researchers[2] but the laws regulating the secondary use of them show variety in conditions and required safeguards[3] in Europe. Some countries are ready for the secondary use of data whilst several barriers exist in other states.[4] The GDPR maintains the traditional purpose limitation principle in the age of bigdata.[5] However, the Regulation is open for repurposing and longer retention periods in the case of statistical and scientific research.[6] Moreover, the GDPR permits the Member States to provide derogations from the rights to access, rectification, restriction and objection of processing in the case of scientific research, which is further elaborated in this paper.

* Postdoctoral Researcher, Academia Sinica, Taiwan, dr.janos.meszaros@gmail.com.
** Assistant Professor/Assistant Research Fellow, Academia Sinica, Taiwan, chihho@gate.sinica.edu.tw.

1 Moerel and Prins (2016), Rouvroy (2016), van der Sloot and van Schendel (2016), Hildebrandt (2013) 7–44.
2 Pascal et al. (2013) 547–60.
3 Anderson (2007) 480–83.
4 Van Velthoven et al. (2016).
5 Zarsky (2017).
6 GDPR Article 5 (e).

The modern healthcare systems and smart devices collect and process a vast amount of data, which may enhance the individuals' healthcare experiences directly and indirectly through scientific research and policy planning.[7] However, it seems impossible, or it would need disproportionate effort to acquire consent from a large number of data subjects for the new processing purposes, which poses complex ethical, legal and technical challenges.[8] To solve these issues, it is crucial to balance the citizens' autonomy, public interest and safeguards when health data is reused for secondary purposes without consent.[9] Since the governments realized the potential of health data registries, de-identification techniques play an increasingly important role in the balancing process. New laws and regulations encourage the application of pseudonymisation, which is the separation of data from the direct identifiers. This de-identification method reduces the risk of the processing and maintains the data utility for research. However, applying de-identification techniques as a legal basis for the secondary use of health data raises concerns as these methods would require further clarification by the laws and industry standards. Moreover, the capability of computer re-identification and the data availability about the citizens is growing, which have a negative impact on the efficiency of these safeguards.

This paper analyses the required public interest and safeguards for the further processing of sensitive data, focusing on the GDPR and its implementations in Germany and England. The article argues that scientific research exemption in the GDPR can be applied with a general level of public interest. In order to elaborate these issues, the article first discusses the research exemption, public interest and de-identification techniques in the GDPR. What follows, is the implementation of the GDPR in Germany and England to shed light on the challenges, which researchers and policy planners are facing in the new era of data protection.

2. RESEARCH EXEMPTION IN THE GDPR

The GDPR is a significant step in the EU to shift from the era of 'small data' to big data with several exemptions for scientific research and statistical purposes. The GDPR encourages innovation and technological developments; thus scientific research has a privileged role in the Regulation with several broad exemptions.[10] It is a common practice in scientific research to process personal data for a purpose which is different from the original one ('secondary use' or 'further processing') to pursue new findings.[11] The GDPR acknowledges 'it is often not possible to fully identify the purpose of personal data processing for scientific research purposes at the time of data collection.'[12] This recognition is crucial, since obtaining consent became more difficult under the Regulation, which must be 'unambiguous' and 'specific' to the processing operation.[13] The GDPR, like the Directive,

[7] Institute of Medicine (2013), Vayena and Tasioulas (2016), Jones et al. (2017).

[8] Burton et al. (2017) 1732–33.

[9] Rumbold and Pierscionek (2017).

[10] The GDPR Recital 157 also highlights that 'by coupling information from registries, researchers can obtain new knowledge of great value with regard to widespread medical conditions such as cardiovascular disease, cancer and depression.'

[11] Auffray et al. (2016).

[12] GDPR Recital 33 and 65.

[13] GDPR Article 4 (11).

forbids the data controllers to process sensitive personal data[14] and as a general rule, researchers may only use sensitive data with the data subject's explicit consent.[15] However, the GDPR intends to ease the restrictions on the processing of sensitive data by explicitly permitting the processing for research purposes. To use this exemption, the data controllers need to apply appropriate safeguards e.g., pseudononymisation, based on the Union or Member State law.[16] According to this wide exemption for research, it would be crucial to clarify and harmonise the definition of scientific research and appropriate safeguards in the EU.

The GDPR defines scientific research in a broad manner, which includes 'technological development and demonstration, fundamental research, applied research and privately funded research' conducted by both public and private entities.[17] Furthermore, the Regulation also promotes the technological and scientific developments by citing the Article 179(1) Treaty on the Functioning of the European[18] for achieving the European Research Area. However, the definition of research is in the Recital part[19] of the GDPR. Thus the Member States may tailor the scope of it, since the overly broad interpretation would lead to cases which are against the goals of the GDPR. For instance, a private genetic testing company may further use the collected data for research purposes, without consent.[20] Furthermore, citizens might have concerns about their sensitive data is shared with public or private organisations.[21]

There are situations when data was not collected for research purposes in the first place e.g., when a smartwatch or phone collects data about the users' exercise habits and heart rate. The reason users allow this type of processing because they want to track their health status and improvement (or decrease) in sport activity. These data can be useful later for research purposes, to find unseen correlations. For example, physical activity may indirectly influence health behaviours such as overeating, smoking, substance abuse, stress management and risk taking.[22] However, it would be impossible or need disproportionate effort to get the consent for the new processing from the data subjects. People might change their smartphones, the patients did not provide their phone number at the doctor. Similarly, when the patients visit their general practitioner, their main goal is to get knowledge and improve their health status. However, their collected data (and tissue) provides valuable information for future research, but reaching them for getting their approval for the

[14] GDPR Article 9 (1).

[15] GDPR Article 9(1)(a).

[16] GDPR Article 9(2)(j).

[17] GDPR Recital 159.

[18] Treaty on the Functioning of the European Union, Article 179 (1).
The Union shall have the objective of strengthening its scientific and technological bases by achieving a European research area in which researchers, scientific knowledge and technology circulate freely, and encouraging it to become more competitive, including in its industry, while promoting all the research activities deemed necessary by virtue of other Chapters of the Treaties.

[19] In the EU law, a recital is part of the text, usually the beginning of the law, which explains the reasons for the provisions, and it is not normative, thus legally not binding. Recitals are usually general statements. The Recital of the GDPR gives guidelines for understanding the normative text and the purposes behind it.

[20] Pormeister (2017) 145.

[21] See Stockdale, Cassell and Ford (2018), Wyatt, Cook and McKevitt (2018), Aitken, Jorre and Pagliari (2016).

[22] Blair, Jacobs and Powell (1985).

secondary purpose would pose a serious burden, if it is possible at all. This can lead controversial scenarios, such as the Google DeepMind case in the UK when the Royal Free Hospital under the National Health Service (NHS) provided personal data of around 1.6 million patients to Google as part of a trial to test an alert, diagnosis and detection system for acute kidney injury. Later the Information Commissioners investigation found several shortcomings in how the data was handled, including that patients were not adequately informed that their data would be used as part of the test.[23]

The GDPR addresses these cases by providing a compatibility test, to strike balance between privacy and public interest. When the secondary processing is not based neither on the data subject's consent or a Union or Member State law, the controller can still further process the personal data, but it needs to perform a purpose compatibility test.[24] This test is a novel tool in the GDPR which helps to identify the crucial aspects of the processing to decide whether the new purpose is compatible with the original one. According to the test, when a controller willing to reuse the data will have to consider:

– any link between the original and new purposes;

– the context in which the personal data have been collected; the sensitivity of the personal data;

– the possible consequences of the intended further processing for data subjects;

– the existence of appropriate safeguards, which may include encryption or pseudonymisation.

The sensitivity of the data is also included amongst the factors to decide whether the data can be further processed. The inclusion of sensitivity among the factors suggests that controllers may be permitted to process sensitive data for secondary purposes.[25] Furthermore, the GDPR considers re-purposing for scientific research as a compatible new purpose.[26]

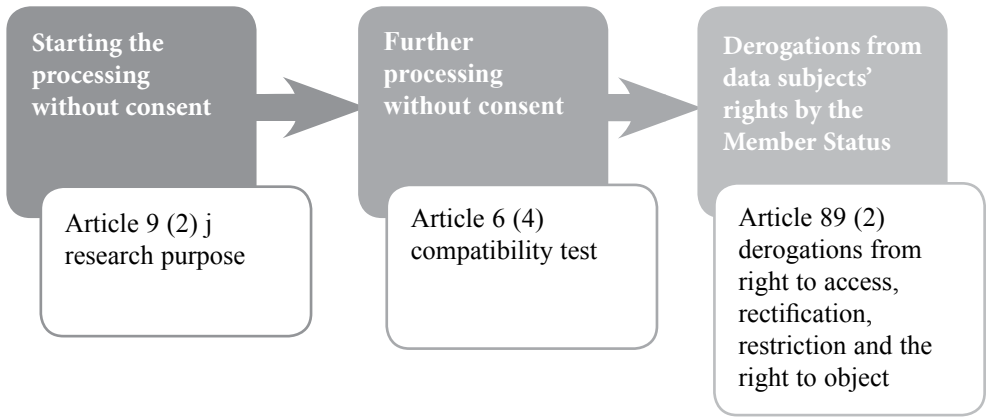

Figure 1. Processing personal data for research purposes without consent in the GDPR

[23] Information Commissioner's Office (ICO) (2017).
[24] GDPR Article 6 (4).
[25] Gabe Maldoff (2016).
[26] GDPR Article 5 (1) b and Recital 50.

The data subjects have several rights to exercise control over their personal data processed for scientific research. During the drafting period of the GDPR, there was hope for reaching harmonisation on this field, but the Regulation avoided to address it comprehensively. An initial reading may suggest that the data subjects have several stronger and new rights in the GDPR, such as the right to be forgotten, portability and object. However, the GDPR allows the member states to decide, if these rights can be applied in the case of scientific research. The main reason for avoiding the harmonisation is the lack of conferred competency of the EU in this field, which is primarily regulated by the member states.[27]

The new rights in the GDPR, such as the right to be forgotten and data portability may also be limited in the case of public interest. For instance, the right to erasure may not apply, if the processing is necessary for reasons of public interest in public health.[28] Similarly, the right to data portability will not apply to a processing which is necessary for the performance of a task carried out in the public interest.[29] When personal data is processed for scientific or historical research purposes or statistical purposes with appropriate safeguards, the data subjects have the right to object to the processing of personal data concerning them unless the processing is necessary for the performance of a task carried out for reasons of public interest.[30]

It remains still unclear how the scientific research exemption will be applied in the corporate context, for product improvement and data analytics. The boundaries are still not clear and the member states have much space left to decide how they would like to apply the exemptions on big data and the repurposing for scientific research. One of the main purposes of the GDPR was to unify the data protection rules in the EU but it is questionable if the Regulation can reach this goal, since many exemptions and important definitions are in the Recital part of the GDPR, which means it is not binding for the member states. This situation may lead to diverse interpretations and forum shopping, which may erode the individuals' privacy, since the member states do not want to fall back in the field of scientific research and potentially lose economical advantages.

3. PUBLIC INTEREST IN THE GDPR

Every society has different expectations for privacy, which makes exceptionally challenging to draft an international data protection law, such as the GDPR. The main goal of the Regulation was to harmonise the data protection law across Europe, but one size cannot fit all – even the GDPR operates with broad opening clauses to leave room for the member states for the further processing of health data for new purposes. Furthermore, the EU does not have enough competency in the fields of scientific research and healthcare to harmonise it comprehensively, which leaves these areas to be primarily regulated by the member states. In general, the data subjects' consent is the primary legal basis for the processing of their health data. Next to receiving explicit consent, further processing may be based on the Union or Member State law. When this is the case, though, it is vital that processing is done for the public interest at least to a minimum extent as this secondary use of data puts the

27 Chassang (2017).
28 GDPR Article 17 (3) (c).
29 GDPR Article 20 (3).
30 GDPR Article 21 (6).

public good above the individual's autonomy. There are numerous data processing scenarios in the GDPR with different levels of public interest. Figure 2 summarises the various degrees of public interest, by attempting to make an order from the perceived lower to the higher level. However, only the implementation, application and enforcement of the GDPR will clarify the meaning of these levels of public interest.

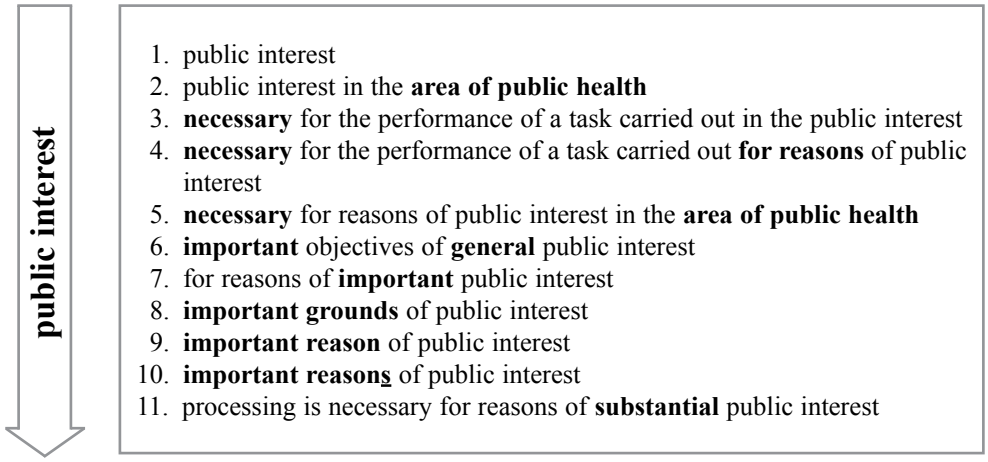

Figure 2. Levels of Public Interest in the GDPR

3.1. General Level of Public Interest in the GDPR

When processing is carried out in the public interest or for the exercise of official authority, the processing needs to have a basis on the Union or Member State law,[31] which specifies the most important aspects of the processing, such as the storage period and purpose limitations. According to this requirement, the Union or Member States laws need to regulate the secondary use of sensitive data for research, particularly the limitations and exemptions of the processing, such as the definition of 'direct care' and 'scientific research'. In general, the processing of sensitive data is prohibited. However, the Union or Member State law may constitute exemptions, where it is in the public interest. The GDPR provides examples of these scenarios, such as health purposes, including public health and the management of health-care services, scientific research or statistical purposes.[32] Furthermore, if the processing for a secondary purpose is necessary for the performance of a task carried out in the public interest, Union or Member State law may determine and specify the tasks and purposes for which the further processing should be regarded as compatible and lawful.[33] Recital 53 also emphasises the importance of public interest by stating that the 'society as a whole' should benefit from the processing of sensitive data for health-related purposes and scientific research.[34] The Recital also requires the processing to be based on Union or Member State law which must meet an objective of the public interest.

[31] GDPR Recital 45.
[32] GDPR Recital 52.
[33] GDPR Recital 50.
[34] GDPR Recital 53.

3.2. Important Public Interest

In the case of important and substantial public interest, the individuals' autonomy may be put aside for the whole society, or even the humanity. The GDPR Recital 46 mentions 'important grounds of public interest' as a basis for the secondary use of health data and provides examples for it, such as humanitarian purposes, including for monitoring epidemics and their spread. The Recital 50 clarifies that the data controller is allowed to process further the personal data irrespective of the compatibility of the purposes for the 'important objectives of general public interest based on Union or Member State law.[35] Processing special categories of personal data may also be permitted for reasons of substantial public interest, on the basis of Union or Member State law, which is proportional and provides appropriate safeguards.[36]

Important or substantial public interest may result in even the deprivation of fundamental rights, though, such as quarantining people[37] means the secondary use of sensitive personal data without de-identification may also be acceptable. However, the balancing process to decide the level of public interest is extremely challenging. For instance, the medical devices need to be safe and reliable, since malfunctions may cost lives thus processing health data for the safety and reliability of these devices has at least general level of public interest. However, the level of public interest in the case of private smart healthcare devices are not as clear, especially if they operate outside the healthcare system. The GDPR permits the processing of the sensitive data if it is 'necessary for reasons of public interest in the area of public health'. The paragraph gives examples for this kind of processing, such as ensuring high standards of quality and safety of healthcare and medicinal products and devices.[38]

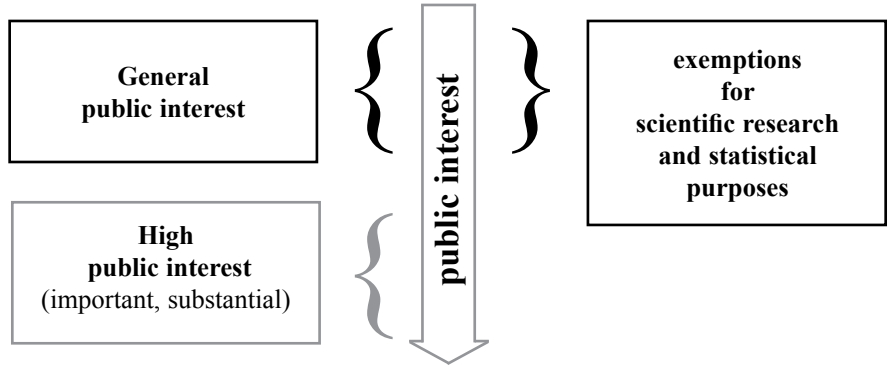

Figure 3. The connection between public interest and the scientific research exemption in the GDPR

Taken together, the recital and the normative text of the GDPR suggest that general level of public interest is sufficient for scientific research purposes. A higher level of public interest e.g., important, substantial, can justify the secondary use of sensitive data itself, the application of research exemptions is not required.

[35] GDPR Recital 50.
[36] GDPR Article 9 (2) (g).
[37] Speakmana, Burris and Coker (2017).
[38] GDPR Article 9 (2) (i).

4. DE-IDENTIFICATION

The de-identification methods represent a broad spectrum of tools and techniques to protect the data subject's privacy. The two ends of this spectrum are clear: personal data without any de-identification, which can directly identify the data subject[39] and the other is the anonymous/aggregated data, which cannot identify the individual.[40] Between these two ends, there is a wide range of methods and techniques, which needs further clarification. Pseudonymization is in the middle of this spectrum, which is the separation of data from the direct identifiers e.g., name, address, ID number, so that re-identification is not possible without additional information (the 'key') that is held separately. The Article 29 Working Party opinion on anonymization characterized pseudonymization more as a security technique. The effectiveness of the pseudonymisation depends on a large number of factors e.g., at which stage it is used, how secure it is against reverse tracing, the size of the database in which the individual is concealed.[41] In general, the strength of the de-identification scales with a loss in data utility and value.

The GDPR provides the Member States the flexibility to draft their regulatory framework for scientific research and secondary use of health data, despite the Regulation lays down the key principles and values. For the secondary use of health data for research, the GDPR requires appropriate safeguards to be implemented by the member states. However, these safeguards are not profoundly detailed, which makes the Regulation both technology neutral and future-proof, but it may fail to harmonize the requirements on the EU level. The specification of these safeguards is typically provided in legislation by the member states, which leaves room for professional codes and further guidance released by the competent data protection authorities.[42] Recognizing the broad spectrum of de-identification techniques and acknowledging them as 'appropriate safeguard' enables the development of regulatory guidance that encourages the maximum use of de-identification and it may open the door for the safe secondary use of data in scientific research. However, the expected level of de-identification would need further harmonization on the EU level.

The European data protection law under the Data Protection Directive has taken a binary approach to the de-identification of personal data, which means the data is either personal data and therefore subject to data protection law or anonymous thus outside of the scope of data protection law. This binary approach may lead to weaker data protection measures. For instance, when personal data is processed for a research purpose that cannot be accomplished with anonymized data, the researchers may have less incentive to use de-identification techniques e.g., pseudonymisation, if they cannot reach any potential benefit from the data protection law e.g., they can further process the data. Even if some level of de-identification is compatible with the purposes of the research and could provide meaningful privacy protections for the individuals, there would be no rewards for the

[39] The GDPR has strict expectation towards anonymisation. Unlike the Health Insurance Portability and Accountability Act (HIPAA) in the United States, which sets forth a rule exempting data from regulation if 18 specific identifiers are removed, the GDPR applies the standard that data is anonymous only when it cannot be identified by any means by any person.

[40] However, anonymised data can still lead to identification of individuals in exceptional cases: Barbaro and Zeller (2017), Narayanan and Shmatikov (2010).

[41] Article 29 Working Party (2007) 19.

[42] Article 29 Working Party (2013) 28.

researchers' efforts. Thus the binary approach can result to weaker data protection.[43] Furthermore, people may be less reluctant to share their sensitive information if they can trust that their data is well protected.[44]

Table 1. The derogation of rights in the case of scientific research purposes

Type of processing (purpose)	General			For scientific research or statistical purposes		
De-identification	anonymous	pseudonymised	Identifiable data	anonymous	pseudonymised	Identifiable data
Right to – access – rectification – restriction of processing – object	⃠	✓	✓	⃠ depending on the Union or Member State law		

The GDPR approaches the de-identification techniques in a more sophisticated way as it differentiates among de-identification techniques and benefits the data controllers for applying them. The anonymisation is still the strongest level of de-identification and anonymous data is not subject to the GDPR. However, the GDPR introduces a de-identification method for the data controllers, namely pseudonymisation, to achieve safer processing and provide more freedom for the further processing of personal data. The Regulation explicitly includes pseudonymisation in both the recital and the normative text – it is the processing of personal data in such a manner that it can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information be kept separately and is subject to technical and organisational measures.[45] There is a wide range of de-identification techniques between anonymous and personal data which needs further clarification. The GDPR clarifies that pseudonymised data is a type of personal data.[46]

However, the definition of pseudonymisation is too broad to know the requirements to reach the adequate level of de-identification. Mike Hintze points out that it is possible to differentiate between two types of pseudonyimised data in the GDPR. In the first case, the data controller does not have the key to re-identify the data subjects. In the second situation, the data controller possesses the key, thus the data subjects can be identified anytime. This differentiation leads to four levels of identifiability in the GDPR:[47]

1. Personal data that directly identifies the data subject,
2. Readily-Identifiable (the controller has the key),
3. Article 11 De-Identified (the controller does not have the key),
4. Anonymous data, it is not possible to identify the data subjects.

[43] Hintze (2018).

[44] Oswald (2014) 245–72, Riordana, Papoutsi and Reed (2015).

[45] GDPR Article 4(5).

[46] However, many scholars are challenging the idea that pseudonymised data constitutes personal data in all cases. For instance: Mourby et al. (2018) 222–33., Bahr and Schlünder (2015) 279–91.

[47] Hintze (2018) 7.

The more nuanced approach for de-identification in the GDPR encourages the controllers to apply the strongest de-identification, which still fits the processing and decreases the privacy risks. Furthermore, the GDPR does not just encourages, it requires the application of the highest level of de-identification for the processing[48] (data minimization principle). This requirement also comes with the 'reward' that the controller does not need to comply with the right of access (Article 15), rectification (Article 16), erasure (Article 17), restriction of such processing (Article 18), notification obligations (Article 19) data portability (Article 20) and the right to object (Article 21) if the controller demonstrates that it is not in a position to identify the data subject. However, in such circumstances, the data subject can provide more information to enable their identification.[49]

The GDPR, when compared to the Directive, provides more room for relying on a legal basis instead of consent for the secondary use of data, if adequate safeguards, especially pseudonymisaion is applied.[50] If the data controller cannot rely on consent and applying anonymization would hamper the purpose of the processing, the GDPR gives incentives to reach the maximum level of de-identification and maintain the data utility for research.

5. THE SECONDARY USE OF HEALTH DATA IN GERMANY

Germany was the first EU Member State which passed its GDPR implementation statue in 2017. The New Federal Data Protection Act[51] (FDPA) came into force on 25 May 2018 and brought significant changes to adjust the national data protection law to the GDPR. The secondary use of data was one of the most debated topics during the drafting period of the FDPA which benefits from the broad, opening clauses of the GDPR by providing numerous exemptions for the secondary use of personal data. The GDPR generally requires consent for the processing of sensitive data. However, the Regulation also permits the Member States to enact laws for processing sensitive data for public health and scientific research, without consent.[52] The FDPA utilizes this permission by containing several provisions applicable exclusively to scientific research. For instance, organizations may process personal data for a purpose other than the original one, for which the data was collected.[53] This rule means that the data controllers in special cases may process the data for a purpose which is incompatible with the original one.[54] This permission is crucial for the entities holding health data, since they can start to process it for a new purpose e.g., scientific research, without the data subjects' re-consent. The FDPA also takes advantage of the

[48] GDPR Article 11 (1) If the purposes for which a controller processes personal data do not or do no longer require the identification of a data subject by the controller, the controller shall not be obliged to maintain, acquire or process additional information in order to identify the data subject for the sole purpose of complying with this Regulation.

[49] Article 29 Working Party (2017).

[50] GDPR Recital 50: 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.

[51] Act to Adapt Data Protection Law to Regulation (EU) 2016/679 and to Implement Directive (EU) 2016/680 of 30 June 2017.

[52] Article 9 (2) i, j.

[53] FDPA 24 (2).

[54] GDPR Article 9 (2).

GDPR's exemption[55] for the secondary use of sensitive data for scientific research by permitting the processing for scientific or historical research, or statistical purposes without consent.[56] However, there are several conditions for this further processing of personal data. Firstly, this processing must be necessary for these purposes and the interests of the controller. Secondly, the controller's interests and purposes need to substantially outweigh data subject's interest in not processing the data. Thirdly, the controller must take appropriate and specific measures to safeguard the interests of the data subject.[57]

The 'old' FDPA also recognised and encouraged pseudonymisation to achieve data minimalization.[58] The FDPA requires the data controllers to make personal data anonymous or pseudonymized as early as possible, in accordance with the purpose of processing.[59]

This approach is flexible as the de-identification method may be tailored for the purposes of the specific data processing scenario. Furthermore, when choosing the appropriate safeguards, the controller may take into account the state of the art technology, the cost of implementation and the nature, scope, context and purposes of the processing as well as the risks of varying likelihood and severity for rights and freedoms of people posed by the processing. The FDPA also benefits from the GDPR Article 89, which permits the Member States to provide derogations from the rights to access, rectification, restriction and objection of processing.[60] If the data controller fulfils the requirements under the FDPA, the data subjects' rights may be limited in connection with the scientific research. For instance, individuals cannot assert rights of access, correction, restriction and objection if it would make the scientific research impossible or cause serious impairment.[61]

In the age of smart healthcare, it is crucial to maintain the connection and feedback among the medical devices, manufacturer and care providers to monitor the performance and safety of the machines and services. However, asking for consent and managing opt-outs may hamper the integrity of the data, which may influence the safety and efficacy of smart healthcare systems and devices. Multiple entities are involved in this process and it would be burdensome to manage consents and apply the withdrawals.[62] The GDPR provides

[55] GDPR Article 9 (2) (j).

[56] FDPA Section 27.

[57] FDPA Section 22 (2) technical organizational measures; increase awareness of staff involved in processing operations; designation of a data protection officer; restrictions on access to personal data within the controller and by processors; pseudonymization and encryption of personal data; confidentiality, integrity, availability and resilience of processing systems; regularly testing, assessing and evaluating the effectiveness of technical and organizational measures.

[58] FDPA § 3 (6a), § 3a.

[59] FDPA Section 71.

[60] GDPR Article 89 (3) 3.Where personal data are processed for archiving purposes in the public interest, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18, 19, 20 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfilment of those purposes.

[61] FDPA Section 27 (2) The rights of data subjects provided in Articles 15, 16, 18 and 21 of Regulation (EU) 2016/679 shall be limited to the extent that these rights are likely to render impossible or seriously impair the achievement of the research or statistical purposes, and such limits are necessary for the fulfilment of the research or statistical purposes.

[62] Felz (2018).

an exemption which helps to solve this issue.[63] If the data processing is necessary for the reasons of public interest in the area of public health, such as high standard of quality and safety of healthcare and medicinal products and devices, the member states may enact laws allowing the use of this data with safeguards.

The German law also takes advantage of this exemption by permitting the processing of sensitive data for the safety and high quality of medical devices and services.[64] However, the exemptions provided by the FDPA for scientific research have their limitations. It is unlikely that the exemptions can be applied to research activities with strong corporate interests e.g., product and service improvement, market research, analytics, etc. Nonetheless, it is not impossible to apply the exemptions on corporate-funded research, if it meets with the strict requirements of scientific research, such as the independence of the researchers and publication of methods and results.

Table 2. The right to opt-out from the secondary use of health data in Germany and England

Country	Germany			England		
De-identification	anonymous	pseudonymised	Identifiable data	anonymous	pseudonymised	Identifiable data
Right to Opt-out	⊘			⊘		✓

The FDPA reaches the expected level of protection of the GDPR but does not go any further. The provided safeguards and the data protection authority in Germany are strong[65] but citizens cannot opt-out from the secondary use of their health data. The FDPA applied all the derogations from the rights provided in the GDPR e.g., right to access, object, which means the citizens cannot exercise these rights in the case of scientific research.

6. THE SECONDARY USE OF HEALTH DATA IN THE UK

The GDPR came into force in May 2018 and the planned date of Brexit is March 2019, which means the GDPR will have been in effect in the UK for almost a year. The new Data Protection Act 2018 (DPA 2018), which implements the GDPR, received the royal assent and became UK law in May 2018. The Act incorporated the GDPR requirements into UK law as the government seeks uninterrupted data flows with Europe and the rest of the world

[63] GDPR 9 (2) (i) processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices, on the basis of Union or Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy.

[64] FDPA Section 22 (1)(1)(c) processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medi-cal devices; in addition to the measures referred to in subsection 2, in particular occupational and criminal law provisions to ensure professional secrecy shall be complied with.

[65] Custers et al. (2017).

post-Brexit. The DPA 2018 also deals with data processing issues that do not fall within EU law e.g. immigration. The DPA 2018 provides more details about health-related research[66] than the German FDPA. However, the DPA 2018 also operates with a broad definition as 'approved medical research' means 'medical research carried out by a person who has the approval to carry out that research from a competent health institution or committee, such as a 'relevant NHS body'. Furthermore, the DPA 2018 requires special conditions – sensitive data can be processed for research purposes if it is in the public interest, with appropriate safeguards and the processing needs to be necessary to reach the aimed purposes of the research.[67]

The DPA 2018 implements the GDPR Article 9(2)(g) with a modified content, which provides a ground for processing sensitive personal data for substantial public interest.[68] The GDPR Article 9(2)(g) requires 'suitable and specific measures to safeguard the fundamental rights and the interests of the data subject' when the sensitive data is processed for substantial public interest. However, the DPA 2018 does not contain these suitable and specific measures connected to this exceptional data processing scenario.[69] The DPA 2018, similar to the German FDPA, also takes advantage of the GDPR Article 89, which allows the member states to have derogations from certain rights e.g., the right to object, when the personal data is processed for scientific research purposes. The DPA 2018 permits the derogations from the following rights to the extent that the application of them would prevent or seriously impair the achievement of the scientific research. These rights are:

a) confirmation of processing,
b) access to data and safeguards for third-country transfers,
c) right to rectification,
d) restriction of processing,
e) objections to processing.

The secondary use of health data is based on a detailed regulation and transparent sharing process in England. The 2006 National Health Service Act[70] allows the Secretary of State for Health to make regulations that bypass the common law duty of confidentiality for defined medical purposes.[71] The Regulations that enable this power are the Health Service (Control of Patient Information) Regulations 2002.[72] The patients can also choose to opt-out, if they want to prevent their data from being shared outside the healthcare system. Furthermore, there is a transition in the opt-out system in England in 2018.

The new system was introduced as a response to the public dissatisfaction with the care.data programme, which aimed to collect patient data from the healthcare system and

[66] DPA 2018 Part 2. Ch. 2. s. 19 (4).

[67] DPA 2018, Schedule 1. Part 1. s. 2 (4).

[68] GDPR Article 9 (2) (g) processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

[69] DPA 2018 Schedule 6, para 12(c) … processing is necessary for reasons of substantial public interest and is authorised by domestic law (see section 10 of the 2018 Act). However, section 10 does not contain suitable and specific measures to safeguard the fundamental rights and the interests of the data subject.

[70] National Health Service Act Section 251.

[71] Smith et al. (2017) 40.

[72] Tassé (2015). and Deloitte (2016) 7–8.

other sources to build a central nationwide database.[73] The data was planned to be accessed by third-party users, including pharmaceutical companies and other private entities,[74] which has raised serious public concern.[75] The programme also aimed to force the General Practitioners (GPs) to share the patient's data, which they highly opposed.

The care.data programme was paused several times because of the criticism from both the public and the members of the healthcare system[76] until it was terminated by NHS England in 2016.[77] The new opt-out system (national data opt-out) is online from May 2018 and similar to the 'old' system's Type 2 opt-out, the national opt-out will apply only where identifiable personal data is shared for purposes beyond individual care[78] for research purposes or managing the health care services.[79] There has been significant changes in the processing of health data in England. However, under both the 'old' and 'new' opt-out systems, the patient's choices are not upheld in the case of direct care and de-identified data.[80] De-identification may involve pseudonymisation and this broad exemption clearly shows the influence of this widespread technique, which has a significant role among the safeguards in the GDPR. Pseudonymisation constitutes a wide exemption so it is crucial to clarify the expected safety levels and techniques by the laws and industry standards, because it varies in the member states.

In England, the patients have the option to withdraw from the secondary processing of their health data. This option is above the requirements of the GDPR, which allows the restriction of the data subjects' rights in the case of scientific research and public health purposes. However, the broad exemptions make the opt-out system less effective.

7. CONCLUSIONS

Modern healthcare devices and applications need a vast amount of data for the constant improvement and safe operation. The most important and reliable sources would be the national registries. This paper has pointed out that there are three main ways in the GDPR to further process personal data for research, without consent: public health purposes, scientific research exemption or important public interest. In these cases, the GDPR permits the member states to derogate from certain rights provided for the data subjects, such as the right to access or object.

[73] Vezyridis and Timmons (2017) 2.

[74] Link 7.

[75] Sterckx and Cockbain (2014) 227–228., Ipsos (2016)

[76] Carter, Laurie and Dixon-Woods (2015) 404.

[77] Link 2.

[78] National Data Guardian for Health and Care, *Review of Data Security, Consent and Opt-Outs,* (2016) 57.
'A clinical, social or public health activity concerned with the prevention, investigation and treatment of illness and the alleviation of suffering of individuals. It includes supporting individuals' ability to function and improve their participation in life and society. It includes the assurance of safe and high quality care and treatment through local audit, the management of untoward or adverse incidents, person satisfaction including measurement of outcomes undertaken by one or more registered and regulated health or social care professionals and their team with whom the individual has a legitimate relationship for their care.'

[79] NHS Digital, *About the national data opt-out* (2017 September) 2.

[80] Link 6.

From the GDPR and its implementation in Germany and England, the following relationship between public interest and de-identification can be drawn up.

1. The scientific research exemptions in the GDPR can be applied in the case of general public interest.

2. The high public interest (important, substantial) can be a legal base itself for the secondary use of sensitive data, the application of exemptions is not necessary.

3. The clarification of the different levels of public interest in the GDPR e.g., general, substantial, important, would be crucial, since they have a significant effect on the data subjects' rights.

4. The definition of scientific research is overly broad in the GDPR thus it remains still unclear how the research exemption will be applied in the corporate context, such as product improvement and data analytics.

5. The required level of pseudonymisation is not clear as it is possible to differentiate between two types of pseudonymisation in the GDPR: whether the controller has the key to identify the data subjects. Further clarification on the EU level would be essential.

The application of big data in the smart healthcare has the potential to save lives and decrease expenses. However, achieving these goals would need the secondary use of sensitive data, which requires a constant balancing between public interest and privacy. The GDPR is a significant step to harmonise this balancing process in the EU. Yet, the further interpretation of public interest and appropriate safeguards would be necessary to provide equal privacy rights and affordable, high quality healthcare across Europe.

LITERATURE

Aitken, Mhairi, Jorre, Jenna de St. and Pagliari, Claudia, 'Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies' (2016) 17(1):73. BMC Med Ethics 1–24.

Tassé, Anne Marie, 'A Comparative Analysis of the Legal and Bioethical Frameworks Governing the Secondary Use of Data for Research Purposes' (2016), Biopreservation and Biobanking 207–16.

Rouvroy, Antoinette, '"Of Data and Men". Fundamental Rights and Freedoms in a World of Big Data' (2016) Council of Europe, Directorate General of Human Rights and Rule of Law 1–38.

Anderson, James, 'Social, ethical and legal barriers to e-health' (2007) 76 Int J Med Inform. 480–83.

Article 29 Working Party, 'Opinion 03/2013 on purpose limitation' (2013) 28.

Article 29 Working Party, 'Opinion 4/2007 on the concept of personal data' (2007) 19.

Article 29 Working Party, 'Opinion on Guidelines on the right to data portability' (2017) 13.

Auffray, Charles, Balling, Rudi, Barroso, Inês and others, 'Making sense of big data in health research: Towards an EU action plan' (2016) 8:71 Genome Medicine 1–13.

Bahr, Anne and Schlünder, Irene, 'Code of practice on secondary use of medical data in European scientific research projects' (2015) 4 International Data Privacy Law 279–91.

Blair, Steven , Jacobs, David and Powell, Kenneth, 'Relationships between exercise or physical activity and other health behaviors' (1985) 100(2) Public health reports 172–80.

Burton, Paul et al., 'Policies and Strategies to Facilitate Secondary Use of Research Data in the Health Sciences' (2017) 46.6 International Journal of Epidemiology, 1732–33.

Carter, Pam, Laurie, Graeme and Dixon-Woods, Mary, 'The social licence for research: why care.data ran into trouble' (2015) Journal of Medical Ethics 404.

Chassang, Gauthier, 'The Impact of the EU General Data Protection Regulation on Scientific Research' (2017) 11:709 Ecancermedicalscience 3–12.

Custers, Bart, Dechesne, Francien, Sears, Alan M. et al., 'A comparison of data protection legislation and policies across the EU' (2017) Computer Law & Security Review 2–12.

Deloitte, 'International review: Secondary use of health and social care data and applicable legislation' (2016) 7–8.

Hildebrandt, Mireille, 'Slaves to Big Data. Or Are We?' (2013) 17 IDP. Revista de Internet, Derecho Y Politica, 7–44.
Information Commissioner's Office, Anonymisation code of practice (2012) 11–18.
Institute of Medicine, 'Best Care at Lower Cost: The Path to Continuously Learning Health Care in America' (2013) Washington D.C. 91–133.
Ipsos, Mori, 'The one-way mirror: Public attitudes to commercial access to health data' (2016) 108.
Jones. Kerina et al., 'The other side of the coin: Harm due to the non-use of health-related data' (2017) 97 Int J Med Inform. 43–50.
M. Speakmana, Elizabeth, Burris B, Scott, Coker, Richard, 'Pandemic legislation in the European Union: Fit for purpose? The need for a systematic comparison of national law' (2017) 121 Health Policy 1021–24.
Oswald, Marion, 'Share and Share Alike: An Examination of Trust, Anonymisation and Data Sharing with Particular Reference to an Exploratory Research Project Investigating Attitudes to Sharing Personal Data with the Public Sector' (2014) 3 SCRIPTed 245–72.
Hintze, Mike, 'Viewing the GDPR Through a De-Identification Lens: A Tool for Clarification, Compliance and Consistency' (2018) International Data Protection Law 86–110.
Moerel, Lokke and Prins, Corien, 'Privacy for the Homo Digitalis: Proposal for a New Regulatory Framework for Data Protection in the Light of Big Data and the Internet of Things' (2016) 84–87.
Mourby, Miranda, Mackey, Elaine, Elliot, Mark et al., 'Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK' (2018) 2 Computer Law and Security Review 222–33.
Narayanan, Arvind and Shmatikov, Vitaly, 'Myths and Fallacies of Personally Identifiable Information' (2010) 53 COMM. ACM 26.
National Data Guardian for Health and Care, Review of Data Security, Consent and Opt-Outs, (2016) 57.
NHS Digital, About the national data opt-out (2017 September) 2.
Pascal, Coorevits, Mats, Sundgren, Klein, Bahr et al., 'Electronic health records: new opportunities for clinical research' (2013) 274(6) J Intern Med. 547–60.
Pormeister, Kärt. 'Genetic data and the research exemption: is the GDPR going too far?' (2017) 2 International Data Privacy Law, 145.
Riordana, Fiona, Papoutsi, Chrysanthi and Reed, Julie, 'Patient and public attitudes towards informed consent models and levels of awareness of Electronic Health Records in the UK' (2015) 84(4) Int J Med Inform. 245–46.
Rumbold, John and Kierscionek, Barbara, 'A critique of the regulation of data science in healthcare research in the European Union' (2017) 18 (1):27 BMC Medical Ethics 6–9.
Smith, Sarah, Sibal, Bharat, Linnane, John et al., 'NHS and public health reorganization in England: health protection and emergency planning, preparedness and response perspective' (2017) 2 Journal of Public Health, 40.
Sterckx, Sigrid and Cockbain, Julian, 'The UK National Health Service's 'innovation agenda': Lessons on commercialization and trust' (2014) 2 Medical Law Review, 227–28.
Stockdale, Jessica, Cassell, Jackie, and Ford, Elizabeth, '"Giving something back": A systematic review and ethical enquiry of public opinions on the use of patient data for research in the United Kingdom and the Republic of Ireland' (2018) 6 Wellcome Open Research 1–23.
van der Sloot, Bart and van Schendel, Sascha, 'Ten Questions for Future Regulation of Big Data: A Comparative and Empirical Legal Study (2016) 7 JIPITEC 112–20
Van Velthoven, Michelle Helena, Mastellos, Nikolaos, Majeed, Azeem et al., 'Feasibility of extracting data from electronic medical records for research: an international comparative study' (2016) 16:90 BMC Medical Informatics and Decision Making 1–9.
Vayena, Effy and Tasioulas, John, 'The Dynamics of Big Data and Human Rights: The Case of Scientific Research' (2016) Philosophical transactions. Series A, Mathematical, physical and engineering sciences 374.2083 1–14.
Vezyridis, Paraskevas and Timmons, Stephen, 'Understanding the care.data conundrum: New information flows for economic growth' (2017) Big Data & Society 2.

Wyatt, David, Cook, Jenny and McKevitt, Christopher, 'Perceptions of the uses of routine general practice data beyond individual care in England: a qualitative study' (2018) 8:e019378 BMJ Open 1–8.

Zarsky, Tal, 'Incompatible: The GDPR in the Age of Big Data' (2017) 4(2) Seton Hall Law Review 1005–08.

LINKS

1. Barbaro, Michael and Zeller Jr., Tom, 'A Face Is Exposed for AOL Searcher' (9 Aug 2006) N.Y. TIMES <http://www.nytimes.com/2006/08/09/technology/09aol.html?_r=0> accessed 15 Aug 2017.
2. Department of Health and Social Care, 'Written statement to Parliament: Review of health and care data security and consent' (2016) <https://www.gov.uk/government/speeches/review-of-health-and-care-data-security-and-consent> accessed 21 Aug 2017.
3. Felz, Daniel, 'An English-Language Primer on Germany's GDPR Implementation Statute: Part 2 of 5' (25 September 2017) <https://www.alstonprivacy.com/english-language-primer-germanys-gdpr-implementation-statute-part-2-5/> accessed 5 May 2018.
4. Maldoff, Gabe, 'How GDPR changes the rules for research' (April 19, 2016) <https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research/> accessed 15 January 15 2017.
5. Information Commissioner's Office (ICO), 'Royal Free - Google DeepMind trial failed to comply with data protection law (03 July 2017) <https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/07/royal-free-google-deepmind-trial-failed-to-comply-with-data-protection-law> accessed 01 May 2018.
6. NHS Digital, 'Your Data Matters to the NHS – patient handout' (2018) <https://digital.nhs.uk/binaries/content/assets/website-assets/services/national-data-opt-out-programme/ndop-patient-handout.pdf> accessed 04 June 2018.
7. NHS Digital, 'DAAG register of approved applications 2011-2014' <https://digital.nhs.uk/services/data-access-request-service-dars/register-of-approved-data-releases/release-register-archive> accessed 12 May 2018.
8. NHS Digital, Your personal information choices (2017) <http://content.digital.nhs.uk/yourinfo> accessed 21 Nov 2017.