Iustum Aequum Salutare
XIX. 2023. 3. • 245–257.

DATA SEGREGATION AND ITS PRIVACY ASPECTS[*]

Dániel Necz
PhD student (PPKE JÁK)

1. Introduction

Our personality reflects how we perceive ourselves and how we are perceived by others in society, whereas privacy appears as our intimate sphere to which we can retreat from the outside world. Throughout history, such concepts were understood differently and even today there are a number of different interpretations in line with the relevant cultural, philosophical or societal context. It is underpinned in this respect, however, that personality and privacy often appear as less and less tangible in our digital age. Our online behavior, habits and daily routine are extensively monitored, analyzed, sold and used by a wide number of companies to increase profits or raise brand visibility. Governments also many times use modern technology to explore our most intimate thoughts for various reasons, including national security purposes or the spread of political propaganda. In this respect, data segregation can be seen as a helpful tool, which splits up huge datasets in an attempt to protect against extensive profiling and analyzation by Bigtech and other companies, as well as by governmental agencies. Data segregation can also create transparency and help data subjects exercise their rights concerning a smaller and more digestible scope of personal data.

Data segregation, however, has its limitations and does not necessarily appear to be most adequate tool in protecting personal privacy. In some cases, data segregation can even be harmful. For example, in case of a recently hospitalized patient, the full scope of the patient's medical status may need to be explored by his physician and the hospital may need to access or share other information concerning the patient for various reasons (for example, insurance purposes or to cooperate in a criminal investigation in case the patient was injured as a result of a crime). Under this scenario, data segregation can be useful for managing access rights within the hospital or for protecting the patient from potential bias (for example by only accessing the patient's insurance status or payment data by the hospital's relevant staff members as necessary).

[*] Supported by the ÚNKP-21-3 New National Excellence Program of the Ministry for Innovation and Technology from the Source of the National Research, Development and Innovation Fund.

In other cases, alternative or additional solutions and techniques can be applied, which better protect privacy and other rights of data subjects, such as various other data security measures used against identity theft and impersonation by cybercriminals. This also means that a hospital, for example, can undertake various data security measures to protect its systems and data stored therein, including backups and different monitoring and protective measures. Nonetheless, data segregation can still be used in combination with such other techniques and solutions to serve as another layer of protection, in case of certain databases storing sensitive information, for example.

In accordance with the above, this paper explores the different interpretations of privacy in a philosophical and historical context and by highlighting the differences between European and American understanding of such concept. The paper also describes different approaches to data segregation, the relevant techniques, as well as some aspects of the relevant European and American regulation. With regard to the above, the paper further describes why data segregation could not be regarded as the best tool for protecting natural persons and privacy in certain cases in the digital age and explores alternatives and various other protective techniques. This paper further highlights the cases, where data segregation – either alone or jointly with other techniques could still be used effectively and explores its effects on data processed for scientific research purposes and on data processing by artificial intelligence (AI).

It is underpinned with respect to the above that this paper does not endeavor to give an exhaustive description of the relevant European and American court and authority practice or an analysis of each relevant legislation, and instead focuses on the main European Union and US legal requirements and concepts. Bearing this in mind, the paper only highlights some relevant court or data protection authority practice to help better understand how privacy can be interpreted, and how data segregation or other data security measures can be effectively used to protect personal privacy.

2. Privacy and Natural Persons

Privacy has long been regarded as a concept related to human personality. It is underpinned in this respect that the essence of human beings and privacy cannot be interpreted without a specific natural person, since neither non-human entities nor inanimate objects or the general public – as a whole – has the right to privacy. The concept of privacy, therefore developed along personhood and how human personality was perceived in various eras. Privacy, humanity and human personality, however, still have various interpretations in different contexts and with respect to different cultural, social and other relevant aspects, therefore we only summarize the main historical aspects on the relations between privacy and human personality.

We further undertake to elaborate on some basic philosophical interpretations and current major issues concerning privacy and to highlight how privacy is interpreted in the European Union and the United States. We do not endeavor, however, to follow an all-encompassing approach and to extensively discuss the philosophical aspects of such concepts under each historical and cultural setting, since such analysis would go beyond the scope of this paper.

2.1. The concept of privacy

Personhood has long been identified and interpreted before scholars started debating on the aspects of personality and privacy. Our ancestors created paintings on the walls of caves depicting successful hunts, family events and tribal life. In almost every culture, people decorated their bodies since the dawn of times, and wore furs, bones or special garments to express their personality, beliefs and social status. In addition to our physical appearance, however, the human intellect has also been regarded as a core characteristic of human beings, which strongly relates to our personality. Bearing this in mind, the Greek philosopher Plato also highlighted rational thinking and nature as an essential human characteristic.[1] Similarly, Aristotle, and following his thoughts, Saint Thomas of Aquinas further emphasized the rational part of the human character which distinguishes humans from other forms of life.[2] Humans, however, could not flourish solely, and create communities and, as highlighted by Thomas Hobbes, political authority.[3] Without such authority, a community could not survive on the long run, therefore human beings subject themselves to decisions of a sovereign or a government that is only restricted to step into the innermost sphere of individuals without the necessary legal authorization.

As highlighted above, the concept of personhood has centered around humans as rational beings, who form a community and live as part thereof. This also means that characteristics, traits and acts of each individual all have a meaning within the community. Certain information appears to be more sensitive and access to such information and its propagation can harm the individual and undermine the trust within the community. In order to protect the innermost sphere of individuals and their family, the concept of privacy emerged and became a fundamental right.

Privacy first appeared as a basic need and as a shelter from strangers and the dangers of the ancient world. Our forefathers created simple dwellings from wood and leaves and made clothes to protect and cover their bodies. They later started to feel ashamed and weak without their clothing and nudity became more and more a subject of moral and religious criticism.

Besides its moral and religious understanding, however, privacy also appeared as something unrelated to the public, especially in Greek philosophy. Aristotle, for example, depicted certain virtuous activities as private as opposite to public activities and also characterized places as private with respect to the nature of actions taking place therein.[4] The concept of privacy later continued to develop more independently from morality and religion along with societal changes. As people did more and more

[1] Robert W. Hall: Plato and Personhood. *The Personalist Forum*, vol. 8., no. 2. (1992) 89.

[2] Miguel García-Valdecasas: Psychology and mind in Aquinas. *History of Psychiatry*, vol. 16, iss. 3. (2005) 292. 10.1177/0957154X05051920. hal-00570823

[3] Garrath Williams: Thomas Hobbes: Moral and Political Philosophy. *Internet Encyclopedia of Philosophy*, (without year) https://iep.utm.edu/hobmoral/

[4] Judith A. Swanson: *The Public and the Private in Aristotle's Political Philosophy.* Cornell University Press, 1992. 2. JSTOR, http://www.jstor.org/stable/10.7591/j.ctvn1t9wp

things inside their houses and less and less in public spaces, more activities became private. The understanding of privacy was further influenced by a number of other cultural, historical and various other factors, such as modern technology and the emergence of civil rights and related social movements.

In modern Europe, privacy became to be understood as an important freedom and also as a basic human right with respect to its earlier denial by totalitarian regimes of the twentieth century. Article 8 of the European Convention of Human Rights expressively includes that "everyone has the right to respect for his private and family life, his home and his correspondence."[5] Such sentence has been considered extremely important after the end of World War II and the fall of fascist regimes. In addition, by the end of communism in Eastern Europe, more and more countries introduced comprehensive privacy laws in order to expand the values of democratic societies and to protect citizens from unlawful processing of their personal data by state and other actors. European regulation by the time, however, rather appeared to be disjointed and contradictory with regard to the fact that member state privacy regulations often provided different levels of protection and requirements, which made it hard for entities undertaking international data transfers to comply. The European Data Protection Directive[6] was a game changer in this respect, since it provided a unified framework for privacy protection in the European Union. The technological developments and social changes, as well as different aspects of member state regulation, however, led to the replacement of the Directive and its repealing by the European General Data Protection Regulation[7] (GDPR), a new comprehensive European privacy regulation, which also became a model law for future regulatory efforts outside of the EU.

In the United States, the concept of privacy has taken a rather different path and generally evolved from the right to be let alone from external interference, including involuntary monitoring, as well as the protection of one's reputation. One of the first major American writings on privacy, "The Right to Privacy" from 1890 by Samuel D. Warren and Louis D. Brandeis also highlighted the importance of privacy and the need for its protection in the wake of technological advancements and the evolving influence of the media.[8] Case law especially intensified in the second half of the 20th century; a gamechanger in this respect was the decision of the U.S. Supreme Court in the case Griswold v. Connecticut in which marital privacy and right of married couples to buy

[5] Article 8 of the European Convention of Human Rights. https://tinyurl.com/368v9cnw

[6] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281, 23.11.1995, p. 31–50.

[7] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), OJ L 119, 4.5.2016, p. 1–88.

[8] Samuel D. Warren – Louis D. Brandeis: The Right to Privacy. *Harvard Law Review*, vol. 4., no. 5. (1890) 196. https://doi.org/10.2307/1321160

contraceptives without undue government restrictions and interference was recognized and highlighted.[9]

Based on earlier court practice and literature, Berkeley law professor, William L. Prosser published the landmark paper "Privacy" in 1960. In this paper, Prosser highlights the kinds of invasion of privacy and related torts, which includes 1) the intrusion into private affairs, 2) public disclosure of embarrassing private facts, 3) false publicity, as well as 4) appropriation of name and likeness.[10] Such elements still remain dominant in the US concept of privacy, however, the influence of European interpretation of privacy and data protection has also left its mark on US legislation in the recent years with respect to the GDPR and its effect on international data transfers. Besides the GDPR serving as a model privacy regulation, the US privacy legislation (and especially state legislation) also highlights the importance of protecting consumers from undue corporate influence, as well as from mass profiling and analysis. State statutes protecting consumer privacy includes the California Consumer Privacy Act (CCPA)[11], as well as a growing number of similar state acts enacted by other states.[12] Initiatives on the level of federal legislation have also been taken, including, for example, the Platform Accountability and Transparency Act[13], aimed at holding certain platform and online service providers more accountable and their data processing activities more transparent.

Although the concepts of privacy took a different path in Europe and in the United States, it is highlighted that under both approaches, the need for enhanced protection of privacy from invasive technology and surveillance practices remains a common focus point.

It is also noted, however, that personhood and personality rights are mentioned more and more often in case of non-humans. This includes the debate about algorithmic or robotic personality and cases where AI solutions are used without adequate human supervision. For example, the US company behind the legal AI solution, DoNotPay was sued recently for unlawful practice of law and for performing legal work below the standard required from legal professionals.[14] Lawyers relying on non-existent cases "found" by the AI solution, ChatGPT, were also subject to sanctioning for not reviewing such cases before including them in their legal brief.[15] These cases also highlight that until a robotic or AI personality is recognized, a legal entity or a natural person developing, using or overseeing AI needs to be held accountable for the AI's actions. And while privacy can certainly only be attributed to human beings, it is important to

[9] Griswold v. Connecticut, 381 U.S. 479 (1965)

[10] William L. Prosser: Privacy. *California Law Review*, vol. 4., issue 3. (1960) 389.

[11] https://tinyurl.com/2vv4e64h

[12] Anokhy Desai: US State Privacy Legislation Tracker. hhttps://tinyurl.com/bd9z22f9

[13] Platform Accountability and Transparency Act. https://tinyurl.com/3ajy23nz

[14] Sara Merken: Lawsuit pits class action firm against 'robot lawyer' DoNotPay. *Reuters,* 2023. 03. 09. https://tinyurl.com/ycy4mbec

[15] Benjamin Weiser – Nate Schweber: The ChatGPT Lawyer Explains Himself. *The New York Times,* 2023. 06. 08. https://tinyurl.com/yc2uewkk

properly regulate AI in order to protect the rights and freedoms of natural persons and to ensure the safe and transparent use of new technologies. This could also help protect personal privacy and ensure that AI is not used contrary to the fundamental interests of human beings and communities.

2.2. Blurred lines in the digital age

In the digital age, the notion of privacy and its relation to human personality is becoming more and more intangible. Users, including children or members of other vulnerable groups many times share sensitive information on themselves and their family members publicly and more and more datasets are compiled and managed online to better serve the digital economy. In order to face such privacy and security challenges threatening personal privacy and other interests of natural persons, various data security measures may be required from companies and state actors alike. Among such solutions, data segregation appears as a versatile technique, which can protect individuals and personal privacy from a wide number of negative effects. It is underpinned, however, that in many cases, it is not an appropriate technique to counter or minimize threats affecting individuals especially with respect to the blurred lines of privacy in the digital age. In case of anonymous profiles, artificial influencers or other online characters, for example, data segregation would in most cases not be regarded as a useful tool since the privacy aspects in case of such profiles or digital personas are quite weak or non-existent (unless they are linked to a natural person). It is further underpinned that data security measures or other protective techniques should only be used in such cases in order to protect law-abiding users but should not shield cybercriminals or other malignant parties.

3. Data Segregation as a tool for protecting privacy

In today’s digital society, data segregation may appear as a viable solution to protect personal data from access by unauthorized parties and to help avoid too much interference into personal privacy or prevent other harmful effects. The separation and restriction of certain information can therefore help protect personal data from hacking, mistaken identification or other incidents and misuse. It is highlighted, however, that data segregation cannot be regarded as a ”Swiss Army Knife” useful in every case. The separation of certain categories of personal data does not always protect against unlawful use, and many times also fails to protect data subjects against certain types of attacks or causes greater harm to privacy than other protective measures.

It is further noted that there are a number of alternatives to data segregation, such as data masking or encryption. Different types of data, business processes or activities affected, as well as different fields of business also require different data security measures. In the following paragraphs, we will highlight the importance of data segregation and summarize cases where implementing other data security measures can be more efficient in protecting personal privacy. In addition, we also endeavor to list the most widely known data security measures, which could serve as an alternative to data segregation or could be used in combination with it.

3.1. Data segregation and its criticism

Data segregation can be regarded as a technique or a group of techniques aimed at dividing data into smaller categories and introducing different access rights to each categories.[16] Besides applying data segregation as a data security measure, the problem of using or accessing too much personal data (especially including sensitive data) can also be countered by data segregation techniques, such as limiting the use of certain personal data for certain, usually pre-determined purposes.

In accordance with the above, the GDPR also sets out basic data security requirements for both controllers and processors of personal data, which are required to "implement appropriate technical and organizational measures" by "taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons"[17]. Such measures can include various types of data security solutions and techniques, such as anonymization or data segregation. It is further underpinned that besides the general requirements of the GDPR, EU member state laws and different sector-specific regulations and best practices may also apply, which have further requirements for certain controllers and processors. For example, banks or health service providers are generally subject to more stringent data security requirements under most jurisdictions.

In the US, data security requirements are generally prescribed by state legislation, sector-specific authority practice or industry best practices, whereas federal laws also rather focus on sector-specific aspects. This also means that different legal requirements may apply with respect to the state in which the entity processing personal data does business or is registered, as well as the relevant industry and other circumstances, such as the extent of the given entity's business and its effect on consumers or other individuals.

With respect to applicable laws, sector-specific regulation, best practices, as well as the circumstances of data processing, data segregation can many times be applied in order to protect personal data against unlawful access. In cases, for example, where payment data of a customer are stored separately from information on transactions undertaken and from other information on habits and related possible analytics, it is less likely that a malevolent third party accessing the different datasets can impersonate the customer or lure him/her into undertaking further transactions or giving additional bank account or card information. Data segregation can also be useful in cases where certain systems or databases are used by multiple members of a company group in different jurisdictions. Under this scenario, data segregation can guarantee that datasets relevant only for certain jurisdictions or company group members may be accessed by authorized personnel only.

The segregation of certain sensitive information (e.g. information on race, ethnicity, financial status, etc.) from other categories of information can also help

[16] Data Segregation, *NordVPN* (without year). https://tinyurl.com/3axz9zp4

[17] Art. 32(1) of the GDPR.

prevent discrimination. For example, in cases, where information on ethnicity, racial background can be processed on loan applicants with respect to the applicable law, the separation of such data from other information (e.g. employment or financial status) can help prevent discrimination and guarantee equal opportunities.

In addition to the above, data segregation can further help isolate problematic datasets or data streams until an internal investigation clarifies the issues involved prior to the conclusion of corporate deals, such as mergers and acquisitions. Although this can be rather challenging in case of datasets actively relied on for regular business, more caution is certainly required in cases where a serious data breach happened in the recent past or where seemingly inadequate data security measures were applied. For example, the Information Commissioner's Office (ICO) in the United Kingdom fined Marriott International Inc. £18,4 million due to its inadequate data security practices and related to a cyberattack on Starwood Hotels and Resorts Worldwide Inc. in 2014, which was later acquired by Marriott.[18]

Even in case of appropriately segregating certain datasets, the use of cross-referenced or apparently separately stored non-personal data can lead to the identification of natural persons under certain circumstances. In its decision from 2022, the ICO highlighted, for example, that a dog's name can identify its handler, even though, a dog's name in itself would not be regarded as personal data.[19] Bearing this in mind, natural persons can be identified by deceased relatives, objects they own or based on companies or different organizations that they can be associated with. The GDPR, for example, does not cover the processing of personal data which concerns legal persons in this respect (e.g. the name of the company or contact details of the company itself).[20] National laws also similarly usually exclude information on companies and other organizations from the scope of privacy protection. The situation is less clear concerning smaller, often one-member entities. In German practice, information on one-member entities, such as financial information, can be regarded as the personal data of the member or other associated person as long as there is a strong personal or financial relationship between the entity and the natural person (e.g. the entity is the sole source of income of its individual owner).[21] This is especially true in cases where the sole member of a one-member entity has unlimited liability for the debts of the entity.[22] Personal data of other individuals can also be regarded as personal data of a given data subject being associated with such individual. An example would be the name of a spouse, cohabitee

[18] Information Commissioner's Annual Report and Financial Statements, 2020-21. *ICO*, HC354, July 2021. 30. https://tinyurl.com/56bbbwaa

[19] Freedom of Information Act 2000 (FOIA) Decision Notice. *ICO*, 2022. 02. 08. 6. https://tinyurl.com/2h7zmy5e

[20] See: Para (14) of the GDPR.

[21] From the yearly report of the Data Protection and Freedom of Information Supervisor of Baden-Württemberg, 2018.10.19. 58. https://tinyurl.com/yck4ch6r ;from the 2021 yearly report of the Data Protection and Freedom of Information Supervisor of Berlin, 124–125. https://www.datenschutz-berlin.de/infothek/publikationen/jahresberichte/

[22] From the yearly report of the Data Protection and Freedom of Information Supervisor of Saxony, 2022. 12. 31. 34–36. https://tinyurl.com/5n89ee45

or partner, which could reveal the sexual orientation of the data subject in the given case.[23]

In addition to the above, data segregation can further meet unexpected challenges in cases, where the categorization or the use of data is highly contextual or where data processing tends to push the boundaries of how we approach personality and its link to privacy and personal data. For example, in the metaverse (i.e. a virtually existing and persistent world accessed by and interacted through means of technology)[24] users generally have a virtual profile within the given virtual world and can also choose to stay anonymous many times. Even in cases, however, where the operator of the metaverse platform does not identify the individual user in front of others, a wide number of information are constantly collected on him/her through sensors and other assets used. Such information can especially include personal data related to the user's behavior, responses, movements, emotional status, as well as other sensitive information, which can create substantial risks to the privacy and wellbeing of users unless privacy requirements are adequately taken into account.[25] Data segregation could generally not protect from such risks, bearing in mind that the segregation of such data would largely hinder the provision of services, therefore other guarantees need to be put in place, such as the strict definition of the purposes of data processing and the processing of only such data, which are necessary for the given purpose.

It is also worth noting that controllers implementing data segregation solutions many times also do not take into account cultural and personal aspects of individuals affected. For example, a record or a dataset including a person's substance usage habits could identify the individual's religious or cultural background in certain situations, bearing in mind that the consumption of certain plant-based substances is many times related to religious or cultural practices followed by certain groups. For example, dried peyote, a cactus grown in Mexico and the United States, is regularly used by Native American religious groups, whereas cannabis is many times used by members of the Rastafari movement. In such cases, jointly using such information (e.g. indicating that substance use is related to a religious practice) or using the part of such information for specific purposes (e.g. for medical examination only) could be more effective in protecting the given individual's personal privacy.[26]

[23] C-184/20. Judgment of the Court (Grand Chamber) of 1 August 2022. OT v Vyriausioji tarnybinės etikos komisija.

[24] Eric Ravenscraft: What Is the Metaverse, Exactly? Everything you never wanted to know about the future of talking about the future. *WIRED*, 2022. 04. 25. https://www.wired.com/story/what-is-the-metaverse/

[25] Christian Ivanov: Metaverse. *European Data Protection Supervisor* (without year). https://tinyurl.com/5eke6hwk

[26] Daniel Necz: No Man is an Island – Data Segregation, Personhood and Privacy. *Harvard Law School LL.M. paper,* 2021. 45–46. (unpublished paper).

3.2. Additional data security measures and alternatives

As highlighted above, data segregation can be regarded as an effective data security measure in some cases, however, it could not be regarded as an appropriate measure in other cases, where separation of certain datasets could not effectively protect the given individual or only exposes him/her to further risks. In such cases other data security measures may be more appropriate and help protect personal privacy more efficiently.

Pseudonymization, for example, is a commonly used data security measure to protect personal data and involves cases where the data subject can only be identified by additional information. If a set of patient data, for example, can only be accessed by providing a code that the patient received, his/her related personal data are pseudonymized, since access to such data is strongly restricted, however, the information protected by pseudonymization can still identify the data subject, therefore such information is considered personal data.[27] The protection afforded by encryption can be similar. In this case, however, data are translated from plaintext to cyphertext, and users need specific keys for encrypting or decrypting the given information.[28] Anonymization, on the other hand, involves cases where the given information no longer relates to an identified or identifiable natural person. Such information is not regarded as personal data and are not subject to protection afforded by the GDPR or other privacy laws.[29] For example, statistical data or similarly aggregated information can be regarded as anonymized, and therefore as non-personal data, unless such data can be related to a natural person. In many cases, however, anonymization techniques can be improperly implemented or become outdated, which can thereby lead to the reidentification or a false sense of protection. Therefore, it is essential for organizations using anonymized datasets to act accountably and to update the datasets they use as necessary.[30]

Data filtering is also a recommended data security measure and involves a number of techniques aimed at removing redundant and often sensitive personal data, and thus to protect the privacy of affected individuals.[31] Under this scenario, large datasets many times collected by platform operators can be filtered, making sure that sensitive information are not shared with any other actors or used for any incompliant purposes that would otherwise not require the use of such information.

It is worth mentioning that the term of 'data masking' is also widely used for techniques aimed at replacing original data with fictious equivalents, and which can also involve anonymization, pseudonymization or other de-identification techniques

[27] See: Para (26) of the GDPR.

[28] What is encryption? Data encryption defined. IBM website: https://www.ibm.com/topics/encryption

[29] See: Para (26) of the GDPR.

[30] Necz (2021) op. cit. 49–51.

[31] Margaret Rouse: Data Filtering. *Techopedia,* 2013. 10. 28. https://www.techopedia.com/definition/26202/data-filtering

and solutions.[32] This can also include the use of synthetic data as an alternative to data segregation. Synthetic data are widely used, for example, for training AI solutions, and are generated from an original dataset by reproducing certain characteristics and structural elements.[33] By using synthetic data or other data masking solutions, the risk of re-identification and invasion of privacy can be minimized without any need to segregate the de-identified data further used.

With regard to the above, the combination of multiple data security measures is generally more efficient in the vast majority of cases than relying on data segregation or any other single technique or solution. Data segregation, therefore, can be a part of a successful data security strategy and not an ultimate solution to better protect personal data. For example, the segregation of sensitive information stored on users by a health monitoring application could help protect against unnecessary use of such data or unauthorized access, however without any additional security measure applied, the segregated dataset would still be vulnerable and not be sufficiently protected from external attacks. Therefore, in cases where a segregated dataset is further encrypted or can only be re-linked to the user by applying a code or a key, which only the user has, the likelihood that data would be compromised is a lot smaller. It is also underpinned, that in cases, where data segregation could impede helping the users or complying with their subject access or other requests, data segregation is not regarded as the appropriate measure. This also means that an organization processing huge datasets often encompassing various types of data needs to have an internal data security policy and a comprehensive data security strategy or program to adequately apply each data security measure, often in combination and with respect to the security requirements and the needs and expectations of the individuals affected.[34]

3.3. How data segregation can help protect the integrity of scientific research?

It is important to note that data segregation cannot be used as a tool to hinder scientific progress, social development or the digital economy. The GDPR also makes it clear that further processing of personal data collected for another purpose shall be compatible with such initial purpose assuming that the further data processing takes place for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.[35] In this respect, personal data can also be stored longer than initially deemed necessary subject to the implementation of appropriate technical and organizational measures.[36] It is noted that such measures may include pseudonymization unless this hinders the purpose of such further processing, and that in cases where the

[32] What is Data Masking? *Delphix Glossary* (without year). https://www.delphix.com/glossary/data-masking

[33] Robert Riemann: Synthetic Data. *European Data Protection Supervisor* (without year). https://tinyurl.com/54wmvy44

[34] Necz (2021) op. cit. 52–53.

[35] Art. 5(1)(b) of the GDPR.

[36] Art. 5(1)(e) of the GDPR.

data subjects do not need to be identified for such purposes, the further processing must be undertaken in that manner without relying on personal data identifying natural persons.[37] Member state laws may also specify requirements concerning data processing for the above referred purposes in accordance with the GDPR.

Bearing the above in mind, it is worth noting that data segregation can also be used as an effective tool to protect the data protection rights of individuals affected by scientific research or another processing specified above. For example, segregated datasets may be relied on in different phases of the research and other data security measures can further be applied to enhance privacy protection.

3.4. What about artificial intelligence?

The use of data segregation techniques can many times also disrupt the effective use of AI solutions, bearing in mind that such solutions often need to rely on huge datasets. Without sufficient data to rely on, AI can be less reliable, and decisions made by AI can be less accountable.[38]

It is also worth noting that often old or outdated datasets are useful for AI development, bearing in mind that developers can compare them to other datasets or find out what led the given solution to an incorrect decision or a non-compliant procedure.[39]

Bearing the above in mind, other solutions may be more effective in protecting personal privacy from dangers and threats posed by AI. This could include – inter alia – risk assessment, human oversight and transparent use of technology. Data segregation in this context could more likely be applied effectively to prevent the use of irrelevant data, and thus to also prevent discrimination or other negative consequences resulting from such data processing.

4. Closing remarks

Privacy is a right and a key concept, which helps protects us as social beings, especially including our intimate relations and innermost sphere. Such intimate sphere, however, remains vulnerable to extensive monitoring and other invasive practices and external attacks, and require the application of effective protective measures, which are often rather technical and complex in the digital age.

With regard to the above, data segregation can be regarded as an effective measure for separating certain data in order to protect the separated datasets or to prevent harmful and unwanted effects on individuals (e.g. discrimination or undue influence).

Data segregation, however, could not be used effectively in a number of scenarios, for example, in cases, where different datasets can still be cross-referenced or linked

[37] Art. 89(1) of the GDPR.

[38] Necz (2021) op. cit. 34.

[39] Necz, Dániel: A mesterséges intelligencia felhasználásával történő adatkezelések egyes sajátos szempontjai (Certain Aspects of Data Processing Using Artificial Intelligence). *Acta Humana,* vol. 10., no. 3. (2022) 103. https://doi.org/10.32566/ah.2022.3.4

together due to inadequately applied data security measures. In these cases, a number of other measures and techniques could be more effective in protecting personal privacy, such as encryption of the relevant communication or different pseudonymization techniques. There is no single solution for protecting personal data in most cases, however, therefore a combination of different technical and organizational measures could generally be more effective in protecting the privacy, as well as other rights and interests of individuals.

It is also highlighted, however, that data segregation should not be an adequate tool in cases where it would overly hinder technological progress, the achievement of other legitimate and rightful purposes and in cases where privacy is less relevant (such as in case of artificial or other online profiles not linked to natural persons, such as users).

In addition to the above, we must also not forget that privacy and its relation to the human personality is constantly reshaping in the digital world. In order to efficiently protect it, we must always understand the relevant technological environment and the effect of technology on individuals.