Anonymisation and Pseudonymisation as Solutions for the Protection of Personal Data

Antonio Landi, Francesco Capparelli, Giulia Finocchiaro, Sofia Pini

This article, focused on the intricacies of anonymisation and pseudonymisation as pivotal techniques for personal data protection within the framework of the General Data Protection Regulation (GDPR), was developed in the context of the PREVENT PCP project which is based on the Pre-Commercial Procurement (PCP) approach. This EU-funded initiative, as a collaborative effort aimed at developing new innovative technologies for early detection and tracking of unattended items in public spaces safeguarding public safety using among others video surveillance techniques, has been instrumental in shaping the discussions and analyses presented herein.

Throughout the development of this article, key insights and lessons learned from the PREVENT PCP project have been integrated, particularly in relation to the practical application and challenges of anonymisation and pseudonymisation in safeguarding personal data. The project's strong link and dependency on innovative data protection solutions have enriched our understanding and approach to these critical topics, emphasising the need for robust and adaptable strategies in the face of evolving technological landscapes and data privacy concerns.

Examples and elements derived from the PREVENT PCP project, including the exploration of state-of-the-art anonymisation and pseudonymisation techniques such as K-anonymity, L-diversity, and the application of differential privacy, reflect the project's influence on our comprehensive analysis. These techniques, crucial for ensuring the privacy and security of personal data in various sectors, have been examined within the context of their relevance and application to the goals of the PREVENT PCP project.

The insights gained from the PREVENT PCP project have not only contributed to a deeper understanding of the current state of data protection but also highlighted the ongoing challenges and opportunities for innovation in the field. As we continue to navigate the complexities of personal data protection, the lessons learned under the auspices of the PREVENT PCP project serve as a valuable foundation for future research and development efforts in this critical area.

Introduction

Anonymisation and pseudonymisation are two key concepts in the field of European Union's data protection laws, in particular the EU Regulation 2016/679 *"on the protection of natural persons with regard to the processing of personal data and on the free movement of such data"* – General Data Protection Regulation (hereinafter "**GDPR**")[1]. In fact, the principles outlined by the GDPR apply exclusively to personal data, which refers to any information concerning an identified or identifiable natural person (i.e., the data subject).
Indeed, while personal data subjected to pseudonymisation – data that could be attributed to an individual through the use of additional information – should be considered information concerning an identifiable natural person, the GDPR rules do not apply to:

- anonymous information, meaning information that does not relate to an identified or identifiable natural person;
- personal data rendered sufficiently anonymous to prevent – or no longer allow – the identification of the data subject.

It is immediately apparent, therefore, how important anonymisation and pseudonymisation are for a data controller aiming to:

- exit the scope of the GDPR, thereby avoiding the need to comply (or continue to comply) with all applicable regulations contained therein (thanks to data anonymisation);
- work to reduce the risks to data subjects while remaining within the scope of the GDPR, as pseudonymization is legally considered an adequate security measure, also in the context of "privacy by design" principle.

1. Anonymisation *vs* Pseudonymisation

The concept of Anonymisation. Anonymisation is the process of rendering personal data anonymous. According to the European Union's data protection laws, in particular the EU Regulation 2016/679 *"on the protection of natural persons with regard to the processing of personal data and on the free movement of such data"* – General Data Protection Regulation (hereinafter "**GDPR**")[2], anonymous data is *"information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject*

[1] 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) http://data.europa.eu/eli/reg/2016/679/2016-05-04.
[2] 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) http://data.europa.eu/eli/reg/2016/679/2016-05-04.

is not or no longer identifiable". Therefore, the concept of "**anonymity**" is closely linked to the identifiability of an individual. Recital 26 of the GDPR specifies that data can be considered anonymous or sufficiently anonymous if, respectively:

- They do not refer to an identified or identifiable natural person, or
- They prevent or no longer allow the identification of the data subject.

The abovementioned recital clarifies that to determine whether a natural person is identifiable, all means reasonably likely to be used by the data controller or a third party to identify, directly or indirectly, the person to whom the data relates should be considered. In order to ascertain the reasonable likelihood of using means to identify the natural person, the same recital continues, the overall set of objective factors should be taken into account, including the costs and time required for identification, taking into account both the technologies available at the time of processing and technological developments.

The concept of Pseudonymisation. Very different from "anonymisation" is the concept of "pseudonymisation", a term defined under Article 4, No. 5, of the GDPR as *"the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person"* (such as logical separation and restricted access). Therefore, in contrast to anonymised data, pseudonymised data can be attributed to the individual to whom it relates, but in order to do so, it is necessary the use of additional information. Pseudonymisation can thus be considered as a measure useful for ensuring security in the processing of personal data, which data controllers should adopt in compliance with the obligations arising from Article 32 of the GDPR; however, it does not exclude the nature of personal data, and, therefore, it does not preclude the application of the relevant data protection law.

In this regard, **datasets** which include personal data may contain direct and indirect identifiers, which allow an individual to be identified or become identifiable. A **direct identifier** is a specific information that references to an individual, such as name or an identification number. An indirect identifier (also called quasi-identifier) is any piece of information (e.g. a geographical position in a certain moment or an opinion about a certain topic) that could be used, either individually or in combination with other quasi-identifiers, by someone that has knowledge about that individual with the purpose of re-identifying an individual in the dataset. The **re-identification likelihood** is the probability in a given dataset of re-identifying an individual, by turning anonymised data back into personal data through the use of data matching or similar techniques. The utility of a dataset is a measure of how useful that information is for the intended purpose.

2.Misunderstandings to be avoided when dealing with anonymisation and pseudonymisation

In April 2021, the Spanish Data Protection Authority (Agencia Española Protecciòn Datos – "**AEPD**") and the European Data Protection Supervisor (hereinafter also "**EDPS**") jointly released a document titled "10 Misunderstandings Related to Anonymisation"[3]. The content of this paper, illustrating the 10 misunderstandings, is hereinafter illustrated.

- **Misunderstanding: "Pseudonymisation is the same as anonymisation".** Pseudonymisation is not the same as anonymisation for the reasons explained in the precedent paragraph.
- **Misunderstanding: "Encryption is anonymisation".** Encryption is not an anonymisation technique, but it can be a powerful pseudonymisation tool. The encryption process employs confidential keys to transform information in a manner that mitigates the risk of misuse, ensuring confidentiality for a defined period. While encryption algorithms are intentionally designed to be reversible for accessibility, referred to as decryption, it is crucial to note that the secret keys used for decryption, constituting the "additional information" (refer to Misunderstanding 1), can render personal data readable and enable identification. In theory, one might consider deleting the encryption key as a means to render encrypted data anonymous; however, this assumption is inaccurate. The erasure or unknown status of the decryption key does not guarantee that encrypted data cannot be decrypted. Numerous factors impact the confidentiality of encrypted data, particularly over the long term. These factors encompass the robustness of the encryption algorithm and key, potential information leaks, implementation intricacies, the volume of encrypted data, and technological advancements (e.g., quantum computing).
- **Misunderstanding: "Anonymisation of data is always possible"**. It is not always possible to lower the re-identification risk below a previously defined threshold whilst retaining a useful dataset for a specific processing. Anonymization represents a process that seeks to strike an appropriate balance between reducing the risk of re-identification and preserving the utility of a dataset for its intended purpose(s). However, in certain contexts or with specific types of data, the re-identification risks cannot be sufficiently mitigated. This scenario may arise when the total number of potential individuals in the dataset (referred to as the "universe of subjects") is too limited, such as in the case of an anonymous dataset containing only 705 members of the European Parliament. Additionally, risks can persist when the data categories are highly distinctive among individuals, allowing for their isolation (e.g., device fingerprints of systems accessing

[3] Agencia Española Protecciòn Datos & European Data Protection Supervisor: "10 Misunderstandings Related to Anonymisation", accessible here: https://edps.europa.eu/data-protection/our-work/publications/papers/aepd-edps-joint-paper-10-misunderstandings-related_en.

a specific website). Another potential case occurs when datasets encompass a substantial number of demographic attributes[4] or location data[5].

- **Misunderstanding: "Anonymisation is forever"**. There is a risk that some anonymisation processes could be reverted in the future. Circumstances might change over time and new technical developments and the availability of additional information might compromise previous anonymisation processes. In particular, computing resources and emerging technologies, or innovative applications of existing technologies, available to potential attackers attempting to re-identify an anonymous dataset evolve over time. Actually, cloud computing offers cost-effective computing capabilities at levels and prices previously unimaginable. Looking ahead, the advent of quantum computers could redefine the concept of what is currently considered "reasonable means."
  Furthermore, the disclosure of additional data over time, such as during a personal data breach, has the potential to establish links between previously anonymous data and identified individuals. The release of records spanning many decades and containing highly sensitive information, such as criminal records, could still significantly harm individuals or their relatives[6].
- **Misunderstanding: "Anonymisation always reduces the probability of re-identification of a dataset to zero"**. The anonymisation process and the way it is implemented will have a direct influence on the likelihood of re-identification risks. A robust anonymisation process seeks to reduce the re-identification risk to a predefined threshold. This threshold is contingent on various factors, including the presence of existing mitigation controls (absent in the context of public disclosure), the potential impact on individuals' privacy in the event of re-identification, the reasons and capabilities of an attacker to re-identify the data[7].
  While achieving 100% anonymisation is the optimal objective from the standpoint of personal data protection, in certain instances, it may not be feasible, and a residual risk of re-identification must be taken into account.
- **Misunderstanding: "Anonymisation is a binary concept that cannot be measured".** It is possible to analyse and measure the degree of anonymisation. The expression "anonymous data" should not be construed as a binary label, suggesting that datasets can be categorized as either anonymous or not. Instead, every record within a dataset carries a probability of being re-identified, based on its susceptibility to being isolated. A robust anonymisation process systematically evaluates the re-identification risk, emphasizing the necessity to manage and control this risk over time.

[4] Rocher, L., Hendrickx, J. M., & De Montjoye, Y. A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. Nature communications, 10(1), 1-9, https://www.nature.com/articles/s41467-019-10933-3.

[5] Xu, F., Tu, Z., Li, Y., Zhang, P., Fu, X., & Jin, D. (2017, April). Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data. In Proceedings of the 26th international conference on world wide web (pp. 1241-1250), https://dl.acm.org/doi/abs/10.1145/3038912.3052620.

[6] Graham, C. (2012). Anonymisation: managing data protection risk code of practice. Information Commissioner's Office https://ico.org.uk/media/1061/anonymisation-code.pdf.

[7] External guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use (2016), https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/external-guidance-implementation-european-medicines-agency-policy-publication-clinical-data_en-0.pdf.

With the exception of particular instances where data is extensively generalised (e.g., a dataset detailing the annual count of website visitors per country), the re-identification risk is never reduced to zero.

- **Misunderstanding: "Anonymisation can be fully automated"**. Automated tools can be used during the anonymisation process, however, given the importance of the context in the overall process assessment, human expert intervention is needed. On the contrary, it is necessary to analyse the original dataset, its intended purposes, the techniques to be applied, and the re-identification risk associated with the resulting data.

  The identification and deletion of direct identifiers, also known as "masking", constitute a crucial aspect of the anonymisation process. However, it is necessary to follow this step with a careful analysis for other sources of (indirect) identification, typically through quasi-identifiers. While direct identifiers are relatively straightforward to identify, indirect identifiers may not always be apparent, and failure to detect them can lead to a reversal of the process, resulting in re-identification and implications for individuals' privacy.

  Automation can play a crucial role in certain stages of the anonymisation process, such as the elimination of direct identifiers or the consistent application of a generalization procedure across a variable. However, achieving a fully automated process that can identify quasi-identifiers in various contexts or determine how to optimize data utility by applying specific techniques to specific variables appears unlikely.

- **Misunderstanding: "Anonymisation makes the data useless".** A proper anonymisation process keeps the data functional for a given purpose. The primary aim of anonymisation is to prevent the identification of individuals within a dataset. However, the application of anonymisation techniques inevitably imposes restrictions on the potential uses of the resulting dataset. For instance, grouping dates of birth into year intervals can decrease the re-identification risk but may simultaneously reduce the utility of the dataset in certain scenarios. This does not imply that anonymous data becomes useless; rather, its utility hinges on the intended purpose and the acceptable re-identification risk.

  Conversely, personal data cannot be permanently stored beyond its original purpose, awaiting potential usefulness for other purposes. Anonymisation serves as a solution for some controllers, allowing them to detach and discard personal data from the dataset while retaining meaningful utility in the remaining dataset (e.g. anonymising access logs of a website by retaining only the access date and accessed page, excluding information about who accessed it).

  The "data minimization" principle requires the controller to assess whether processing personal data is necessary for a specific purpose or if the same purpose can be achieved with anonymous data. In some instances, this evaluation may lead to the conclusion that rendering the data anonymous does not align with the intended purpose. In such cases, the controller must decide between processing personal data (employing techniques like pseudonymisation) and apply the GDPR, or not to process the data at all.

- **Misunderstanding: "Following an anonymisation process that others used successfully will lead our organisation to equivalent results"**. Anonymisation processes need to be tailored to the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons. Anonymisation cannot be universally applied like a standardized recipe, as the context (including the nature, scope, and purposes of data processing) is likely to vary from one situation or organization to another. The effectiveness of an anonymisation process, measured by re-identification risk, may differ significantly based on factors such as the limited number of recipients versus making data available to the general public.
  Datasets, existing in diverse contexts, may present distinct challenges. Cross-referencing these datasets with anonymous data can impact the re-identification risk. For instance, in Sweden, taxpayers' personal data is publicly accessible, whereas in Spain, it is not. Consequently, even if datasets containing information about Spanish and Swedish citizens undergo anonymisation using the same procedure, the associated re-identification risks may vary.
- **Misunderstanding: "There is no risk and no interest in finding out to whom this data refers to".** Personal data has a value in itself, for the individuals themselves and for third parties. Re-identification of an individual could have a serious impact for his rights and freedoms.
  Attacks against anonymization can manifest as deliberate re-identification attempts, unintended efforts at re-identification, data breaches, or the unauthorized release of data to the public[8]. The first category involves intentional endeavors to re-identify individuals, while the second encompasses scenarios where re-identification may occur inadvertently. The possibility of someone re-identifying at least one person in a dataset, driven by curiosity, chance, or specific interests like scientific research, journalism, or criminal activities, cannot be disregarded[9].
  Assessing the impact of re-identification on an individual's private life can be challenging as it invariably depends on the context and the correlated information. For instance, re-identifying a data subject based on seemingly innocuous movie preferences may lead to inferences about that person's political leanings or sexual orientation. Such particularly sensitive data, however, receive special protection under the GDPR.

3.ICO's Guidelines on anonymisation and pseudonymisation

Also, the English Information Commissioner's Office ("**ICO**") released, on 2021, its guidelines in order to raise awareness in the market regarding the topics on anonymisation and pseudonymisation. In particular, in May 2021, it published the first part ("Introduction to anonymisation"[10]),

[8] KHALED EL EMAM and LUK ARBUCKLE, Anonymizing Health Data (p. 29-33).
[9] KHALED EL EMAM, ELIZABETH JONKER, LUK ARBUCKLE, BRADLEY MALIN, "A Systematic Review of Re-Identification Attacks on Health Data", 11 December 2011.
[10] Information Commisioner's Office, "Introduction to anonymisation", https://ico.org.uk/media/about-the-ico/consultations/2619862/anonymisation-intro-and-first-chapter.pdf.

and in October, the second part ("How do we ensure anonymisation is effective?"[11]) of a work that underwent public consultation, titled: "Anonymisation, pseudonymisation and privacy enhancing technologies guidance". In this work, ICO underlined that data are the lifeblood of the digital economy, and data sharing is key to opening up new opportunities. It recognized the benefits that data sharing can bring to organizations, individuals, and society as a whole, but acknowledge the associated risks.

The abovementioned guidelines complement ICO's data sharing code of practice, which offers practical advice on how to share personal data in accordance with data protection laws. Focusing on anonymization ICO indicated that it is an alternative approach to using or sharing data by ensuring that individuals remain non-identifiable. In addition, effective anonymization techniques provide a privacy-friendly alternative to sharing personal data. However, It is essential to have a reasonable degree of confidence that disclosing or sharing seemingly anonymous information will not result in an inappropriate disclosure of personal data, for example, through "re-identification".

According to ICO, determining the status of information in various circumstances is a key challenge. For instance, you may possess information that is clearly personal data, but its status when processed by another organization or by the public at large may be unclear.

Therefore, similar to the European Data Protection Supervisor (EDPS) and the Spanish Data Protection Agency (AEPD) approach, analysed in the previous paragraph, the ICO emphasizes the need to assess each anonymization case as a separate situation. Although "absolute" anonymization (100%) is desirable, it is not always practical, especially considering the rapid evolution of technology. However, the potential persistence of re-identification risk does not imply absolute ineffectiveness of the anonymization technique; data protection laws do not require anonymization to be entirely risk-free. The specific risk of re-identification must be mitigated so that the occurrence of the event is sufficiently remote.

Following ICO's guidelines, the key factors for the identification of an individual include:

- Identification of a data subject within a dataset by the data controller or another entity ("***singling out***").
- Linking different pieces of information – contained in one or more databases – about the same individual or groups of individuals ("***linkability***"); due to the so-called "mosaic effect," individual data do not provide any information, but when combined with others, they reveal a meaningful picture.
- The possibility of deducing, guessing, or predicting details about someone based on other available information ("***inferences***"); such predictions can also result from analysis processes aimed at finding correlations between different datasets and using them to categorize and profile individuals.

[11] Information Commisioner's Office, "How do we ensure anonymisation is effective?", https://ico.org.uk/media/about-the-ico/documents/4018606/chapter-2-anonymisation-draft.pdf.

Effective anonymization techniques aim to reduce the possible occurrence of the aforementioned identifiability conditions, following a risk assessment based on various factors, including the state of the art of technology. The rapid evolution of technology requires that the mentioned assessment be periodically conducted to evaluate whether the measures adopted at the initial stage (T0) remain valid in subsequent stages (T1, T2, etc.), or if new or different measures are necessary for data to remain anonymous.
Another criterion suggested by the ICO in this context is the so-called "***motivated intruder test***", to assess whether a potential intruder would be able to make identifiable those data subjects whose data are anonymised, if they intended to do so, thanks to additional information in their possession or otherwise accessible/acquirable. The level of competence of the potential intruder, within the test, should also be parameterised to the type of data involved: the presence, for example, of financial, biometric, or otherwise highly confidential data should lean towards the use of reinforced security measures.
Regarding pseudonymization, the ICO warns of the risk that, with reference to a specific dataset, the data controller may consider it "anonymised" despite containing personal data, but in pseudonymized form. In such a situation, the mistaken belief that the GDPR (or other regulations) does not apply – as mentioned earlier, only anonymous data fall outside the scope of the GDPR – could have prejudicial consequences for both the data controller and the data subjects.
The ICO also mentioned some of the advantages of pseudonymization, summarized as follows:

- As envisaged by recital 29 of the GDPR (as well as the so-called "UK GDPR"), pseudonymisation measures are encouraged not only as a security measure but also as a possible tool for general data analysis.
- Pseudonymisation is one of the factors to consider if a data controller decides to continue processing data for a new purpose compatible with the original one.
- Pseudonymisation is a key security measure, both in the design phase of processing and during the implementation of any project.
- Pseudonymization techniques can reduce the risk of harm to data subjects in the event of a data breach and can also make it easier to manage data subject rights (some may not apply if the controller can demonstrate an inability to identify the data subjects).

Therefore, according to ICO's indications, anonymising or pseudonymising personal data can bring significant benefits to a data controller, both in economic and legal terms, provided that the correct measures are implemented, in line with market best practices and the guidance provided by data protection authorities.

4.Deloitte's judgement: what's change?!

With the judgment of April 26, 2023[12] (hereinafter referred to as the "**Judgment**"), the Court of Justice of the European Union (hereinafter the "**Court**" or "**CJEU**") has expressed its views on the concepts of "pseudonymisation" and "anonymisation" in the context of data processing involving primarily two different organizations. One organization acts as the data controller and sender of a set of data, while the other acts as the recipient of such data.

The CJEU has provided an interpretation that it appears – for reasons that will be explained below – appears to be innovative, sparking a lively debate among industry professionals. This debate is undoubtedly fueled by the crucial importance that these concepts hold for data protection law.

As already indicated, establishing whether a piece of data qualifies as personal or not is pivotal for the (non)application of the GDPR. Again, it is worth noting that the GDPR does not apply to the processing of anonymous information, while it fully applies to the processing of pseudonymised data – with all the ensuing implications.

The Case

The case examined by the Court primarily involves the **Single Resolution Board** ("**SRB**"), namely the European Union authority for resolving banking crises, with the mission of ensuring the orderly resolution of troubled banks and **Deloitte**, involved as an independent assessor tasked by the SRB to conduct evaluations. The evaluations aim to ascertain whether shareholders and creditors of Banco Popular would have received better treatment if the bank had undergone a regular insolvency procedure.

In summary, and for our purposes, the SRB, as part of its institutional activities, collected information related to shareholders and creditors who participated in the procedure concerning the right to be heard (hereinafter referred to as the "**Interested Parties**"). This information included their identifying data (necessary for auditing purposes) and their personal observations on aspects relevant to the aforementioned evaluations. Subsequently, the SRB pseudonymised the received observations and transmitted them to Deloitte without providing additional information needed for re-identification. Therefore, the SRB remained the sole entity capable of linking the observations to the data allowing the identification of the authors.

Additionally, during data collection, the SRB informed the Interested Parties about the processing of their personal data but did not disclose that the observations would be transmitted to Deloitte. This informational gap, according to five claiming Interested Parties, violated the obligation to inform them about the recipients of the processed personal data, leading to the proceedings resulting in the discussed Judgment.

[12] Judgment of the General Court (Eighth Chamber, Extended Composition) of the Court of Justice of the European Union , 26 April 2026, https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX:62020TJ0557#t-ECR_62020TJ0557_IT_01-E0001.

Key Points of the Judgment

Having established that the observations and viewpoints of the Interested Parties fall within the scope of "personal data" as per the GDPR, and given that Deloitte did not have access to information enabling the identification of the authors, the Court questioned whether Deloitte should be considered a "recipient" and thus whether the information transmitted to it constituted, in its regard, "personal data" – information related to an identified or identifiable natural person.
Essentially, without specifically addressing the concrete case, the Court criticizes the decision of the European Data Protection Supervisor (EDPS), who had pronounced on the matter following the received complaints. The EDPS concluded, by presumption, that the observations transmitted to Deloitte could be classified as "personal data" solely because the SRB, the data controller and a separate entity from Deloitte, possessed additional information to link these observations to their respective authors.
The Court's position can be effectively summarized in points 96 and 97 of the Judgment, which are partially reproduced below:

- 96. *"It is true that"* – here the Court agrees with the EDPS and refers to the so-called *Breyer* judgment of the Court of Justice of the EU of October 19, 2016, case C 582/14 – *"[...] the fact that the additional information necessary to identify the authors of the comments received [...] was held not by Deloitte, but by the SRB, does not appear such as to exclude a priori that the information transmitted to Deloitte constituted, for Deloitte, personal data"*.
- 97. *"However"* – the Court continues, relying on the aforementioned *Breyer* judgment – *"[...] to determine whether the information transmitted to Deloitte constituted personal data, it is necessary to put oneself in Deloitte's position in order to determine whether the information transmitted to it relates to 'identifiable persons'"*.

In line with point 96, the absence of possession of additional information by the recipient could exclude that the transmitted information (even if held by another entity) constitutes personal data for the recipient.

The Impact of the Judgment

In other words, the position of the Court can be summarized as follows: data that are considered personal by the sender do not necessarily have to be considered as such by the receiving party. Consequently, if the data are not considered personal by the recipient, the data controller is not obligated to inform the data subjects that their data will be processed by the receiving entity as the recipient.
It appears that the Court introduces a subjective relativistic exception to the concepts of anonymisation and pseudonymisation. It clarifies that:

- It is necessary to consider Deloitte's perspective to determine if the information transmitted to Deloitte constitutes personal data.

- The fact that the SRB holds additional information is not sufficient to a priori exclude that the information transmitted to Deloitte constitutes personal data for Deloitte.

The above seems not to align perfectly with a literal interpretation of the relevant rules, for the reasons hereinafter illustrated.
According to recital 26 of the GDPR, it would seem that the anonymity of data should be evaluated in relation to the specific subject involved in a particular phase of processing rather than as a whole. To clarify, if, within an activity, data processed could be attributed by any of the different entities involved in the processing to the person to whom the data refers, then that data in that specific context should be considered personal data. Supporting this interpretation are at least two literal elements of Recital 26: *"To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly"*. The European legislator does not seem concerned about which entity can identify the person; it suffices that at least one entity can do so. Similarly, the methods by which identification is carried out, whether direct or indirect, do not seem to matter. The use of the term "indirectly" suggests, for example, cases where the data controller is able to identify the person through the data processor, and vice versa.
According to this interpretation, as advocated by the EDPS, for the identifiability of a person from a pseudonymised dataset, it would not matter which entity holds the additional information; it would be sufficient that at least one entity holds it.
Regarding pseudonymisation, the mentioned rules do not seem to admit the possibility that data that is pseudonymised in a certain context can be considered anonymous in the same context, depending on the entity processing it. This is evident from the definition of "pseudonymisation" itself, as provided earlier, which introduces, as a condition for data processing to be considered pseudonymised, the fact that *"additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person"*. In other words, the separate storage of additional information and the security measures taken to protect such information (including, naturally, limiting access to them) are indispensable elements for a dataset to be considered pseudonymised. It is surprising, then, that not making the additional information accessible to the data controller could theoretically be considered to render that data anonymous, hence non-personal.
In conclusion, far from hastily interpreting the judgment, in any case, it will be crucial, to qualify the transmitted data as anonymous from the recipient's perspective, to conduct a careful and well-documented assessment of the identifiability level of the data subjects by the recipient. In this regard, contractual agreements between the parties will also be of particular importance, aiming to minimize (almost nullifying, in this case) the risk of re-identification by the recipient, including the risk of the recipient gaining knowledge of the additional information necessary for re-identification.
Once again, in these cases the data controller shall follow the principles of accountability and a risk-based approach.

5.Techniques of anonymisation and pseudonymisation according to the state of the art

In the challenging and rapidly evolving world of data privacy, various techniques have been developed to achieve data anonymisation and pseudonymisation. These techniques are tailored to protect sensitive information, especially in an era where data sharing and processing is ubiquitous in various sectors such as healthcare, finance and academic research.

Anonymisation techniques

Among anonymisation techniques, models such as K-anonymity and L-diversity have emerged as important solutions that address the challenges of maintaining privacy while preserving the utility of data.
K-anonymity, for example, was developed as a sophisticated response to the shortcomings of traditional anonymisation methods, which focused primarily on removing explicit identifiers such as names or identification numbers. However, this method fell short because it did not take into account the possibility of re-identifying individuals by combining other attributes of the dataset with external information. K-anonymity fills this gap by ensuring that each individual in a dataset is indistinguishable from at least “k-1” others based on certain shared attributes. This approach is particularly effective in contexts where data sharing is prevalent, such as academic research or statistical analysis.
To better understand the K-anonymity model, it is important to understand its approach to handling different types of identifiers. The model distinguishes between explicit identifiers, which are direct and unique data points associated with an individual, and quasi-identifiers. These quasi-identifiers are more complex; they may appear harmless in isolation but can lead to identification when combined. K-anonymity addresses this challenge through two primary strategies: generalisation and suppression. Generalisation involves reducing the precision of quasi-identifiers, such as broadening age ranges or generalising geographical locations, while suppression involves removing data points that are too revealing. These strategies are critical to achieving the primary goal of K-anonymity: to conceal individual identities while preserving the overall utility of the dataset.
However, K-anonymity is not without its limitations. A key challenge is to find the right balance between preserving privacy and maintaining the utility of the data. Over-generalisation can lead to a loss of information value, while under-generalisation can make data vulnerable to privacy breaches. In addition, K-anonymity does not fully address the distribution of sensitive attributes within anonymised groups, which can lead to potential information leaks. This requires continuous evaluation and adjustment of generalisation and suppression levels to optimise the trade-off between data utility and privacy.

To improve the effectiveness of K-anonymity, the L-diversity model has been developed. This model complements K-anonymity by focusing on the diversity of sensitive attributes within equivalence classes, ensuring that each class contains a variety of sensitive attributes. This approach significantly strengthens the protection of the dataset against sophisticated inference attacks.
Building on L-diversity, recursive (c, l)-diversity introduces additional parameters to optimise representation within equivalence classes. This model addresses the homogeneity problem prevalent in K-anonymity by preventing a single sensitive attribute from dominating an equivalence class. Although more computationally demanding, this advanced variant provides a tailored privacy solution, particularly useful in datasets where the distribution of sensitive attributes is complex.
The harmonisation of K-anonymity and L-diversity creates an enhanced privacy framework that exploits the strengths of both models for enhanced privacy. However, this integration increases the computational burden and requires careful calibration to maintain the analytical utility of the data.
Looking to the future, emerging technologies such as homomorphic encryption, federated learning, AI-driven anonymisation frameworks, and the incorporation of differential privacy are revolutionising the field of data anonymisation. These advances enhance K-anonymity and L-diversity, providing more sophisticated and flexible solutions to the dynamic challenges of privacy.

Pseudonymisation techniques

Pseudonymisation techniques can be described as a complex and adaptive process involving a mixture of methods, each of which contributes to the overall effectiveness of protecting sensitive data while retaining its utility for different applications.
Data masking, a key technique within pseudonymisation, involves modifying the original data in such a way that individual information is unreadable. However, this alteration is designed to preserve the overall structure and integrity of the dataset. Data masking can be dynamic, where changes are made in real time at the application layer without altering the underlying data in the database, or static, where permanent changes are made at the data source. Dynamic Data Masking (DDM) is particularly useful in scenarios where real-time data interaction is required, while Static Data Masking (SDM) is typically used in non-production environments such as development or testing. The selective nature of data masking allows different data elements to be masked to varying degrees, depending on their sensitivity and the specific requirements of the data.
Tokenisation, another key aspect of pseudonymisation, takes the concept of data substitution a step further. It involves replacing sensitive data elements with non-sensitive equivalents, known as tokens, which have no intrinsic value but can be mapped back to the original data under controlled conditions. The relationship between the token and the original data is maintained in a highly secure, encrypted database known as

the token vault. The effectiveness of tokenisation depends on the robust management of this token vault, which ensures the secure mapping between tokens and original data.

Synthetic data generation is an innovative and forward-thinking approach to pseudonymisation. Instead of modifying existing data sets, this method creates entirely new data sets from scratch. Algorithms analyse patterns, relationships and structures in the original data, capturing its essence without relying on specific individual data points. These algorithms, often based on advanced statistical modelling and sometimes machine learning techniques, then generate new data sets that reflect the patterns and structures of the original, but do not contain any actual individual data points. This method is particularly useful in situations where the highest levels of privacy are required, as the synthetic data, devoid of real data points, is almost impossible to trace back to real individuals.

Adding random noise to data sets is a strategy based on the concept of Differential Privacy. It involves mathematically altering the data to protect individual privacy while ensuring that the overall utility of the dataset remains intact. The application of this technique varies depending on the nature of the data. For continuous data, random values, often derived from a Gaussian or normal distribution, are added to each data point. This introduces variation that obscures individual data entries, making it difficult for unintended entities to identify exact values, while preserving the overall trends and patterns in the data set. In the case of categorical or discrete data, quantisation noise is applied by segmenting the data into pre-defined ranges or categories and introducing specific types of noise within these segments. This segmentation and subsequent noise introduction ensures that the nuances of categorical data are preserved, allowing general trends and patterns to be identified without compromising the accuracy of individual data points.

As the privacy landscape continues to evolve, the introduction of new pseudonymisation and anonymisation techniques represents a significant step forward in data protection. However, the implementation of these techniques is not without its challenges: the computational requirements are significant, especially for large datasets. Furthermore, the effectiveness of these techniques is not consistent across all scenarios; it varies greatly depending on the specific characteristics of the dataset, the environment in which the data is used, and how these methods are integrated with emerging technologies. This variability requires a continuous process of adaptation and a full understanding of the strengths and limitations of each technique.

6. Conclusions

In conclusion, the discussion around anonymisation and pseudonymisation in data protection reveals a complex and evolving landscape, closely linked to the nuances of GDPR compliance and technological advances.

Anonymisation, as a method of rendering data unidentifiable, effectively places data outside the scope of the GDPR. However, it's important to recognise that this process is not a fixed state. It requires continuous monitoring and adaptation to maintain its effectiveness against evolving

technological capabilities and re-identification risks. The belief in permanent or absolute anonymisation is a misconception; rather, it is a dynamic state that is constantly influenced by new technologies and changing contexts.

Pseudonymisation, on the other hand, while different in its approach as it still retains links to personal data, remains within the GDPR domain. It serves as a crucial technique for risk mitigation, in line with the principles of “privacy by design”. This method emphasises the importance of proactive and preventative measures in the processing of personal data, taking into account both privacy protection and data usability.

The 2023 Deloitte judgment illustrates the fluid nature of these concepts, particularly in the context of data transfers between entities. This judgment introduces a nuanced understanding of personal data, emphasising a context-specific approach to determining the nature of data and the responsibilities of data controllers. This case highlights the need for a tailored interpretation of anonymisation and pseudonymisation in different scenarios.

In terms of technical evolution, the development of anonymisation and pseudonymisation techniques reflects a continuous effort to balance data utility with privacy protection. Advanced techniques such as K-anonymity, L-diversity and differential privacy represent this ongoing effort. These techniques, together with emerging technologies such as homomorphic encryption and federated learning, represent the cutting edge of privacy strategies. However, the implementation of these techniques is complex, with significant computational requirements and varying effectiveness in different data scenarios. Their application requires expert knowledge, continuous risk assessment and a deep understanding of their potential and limitations.

At its core, the future of privacy lies in a nuanced and adaptive approach to anonymisation and pseudonymisation. The convergence of legal interpretation, authoritative guidance and technological innovation is shaping the current paradigm. As a result, organisations must not only navigate, but actively engage with the multifaceted aspects of privacy to ensure that personal data is protected in accordance with evolving legal standards and societal expectations.