Doctrine

Consultation on Guidelines 01/2025 on Pseudonymisation - Doctrine

Pseudonymisation is crucial for maintaining several core principles of data privacy, as it serves as a measure that meets the minimization principle, acts as a risk-adapted security measure, and provides additional safeguards for data transfers. While the importance of pseudonymisation is indisputable — and we welcome these guidelines in an era of fast-moving technological innovation — they fall short in clarifying how personal data should be defined in the context of pseudonymisation. This lack of clarification presents a significant practical challenge, especially amid ongoing debates in case law and legal doctrine regarding whether personal data should be interpreted in relative or absolute terms. Understanding what is and what is not personal data is key to properly applying and enforcing the GDPR. The stakes are high: if no personal data is processed, the GDPR does not apply.

Doctrine's contribution to these guidelines will focus on the issue of qualifying pseudonymized data with regard to the GDPR.

1. Relative approach

First and foremost, the guidelines provide an unequivocal definition of pseudonymized data, which runs contrary to the case law of the CJUE.

> 22. *Pseudonymised data, which could be attributed to a natural person by the use of additional information, is to be considered information on an identifiable natural person, and is therefore personal. This statement also holds true if pseudonymised data and additional information are not in the hands of the same person. If pseudonymised data and additional information could be combined having regard to the means reasonably likely to be used by the controller or by another person, then the pseudonymised data is personal. Even if all additional information retained by the pseudonymising controller has been erased, the pseudonymised data becomes anonymous only if the conditions for anonymity are met.*

Indeed, previous decisions by the CJUE have tended to favor a relative approach, based on recital 26 of the GDPR:

- CJEU, 19 October 2016, C-582/14, Patrick Breyer v. Bundesrepublik Deutschland on IP addresses [1];

[1] 49. Having regard to all the foregoing considerations, the answer to the first question is that Article 2(a) of Directive 95/46 must be interpreted as meaning that a dynamic IP address registered by an online media services provider when a person accesses a website that the provider makes accessible to the public constitutes personal data within the meaning of that provision, in relation to that provider, where the latter has the legal means which

- CJEU, 9 November 2023, C-319/22, Gesamtverband Autoteile-Handel v. Scania CV AB concerning the identification number of a vehicle in a database [2];
- General Court, 26 April 2023, T-557/20[3] and more recently, Advocate General Spielmann's Opinion in C-413/23 P case, European Data Protection Supervisor v. Single Resolution Board, on the transmission of comments pseudonymised by an alphanumeric code to an audit firm :

> 51. I infer from the wording of those provisions that pseudonymisation leaves open the possibility that the data subjects may not be identifiable, otherwise the wording of recital 16 of that regulation would be pointless. I would add that the final sentences of that recital concerning anonymisation confirm this interpretation: they exclude anonymised data (or data rendered anonymous) from the scope of Regulation 2018/1725, **but exclude pseudonymised data from it only in so far as the data subjects are not identifiable**. If it is impossible to identify those data subjects, they are therefore legally considered to be sufficiently protected by the pseudonymisation process, **notwithstanding the fact that the additional identification data have not been completely erased.**
>
> 57. Thus, it is only where the risk of identification is non-existent or insignificant that data can legally escape classification as 'personal data'.

That would be the case if the identification of the data subject was prohibited by law or practically impossible, for instance on account of the fact that it requires a disproportionate effort in terms of time, cost and manpower (par. 46 - Breyer Case).

In the light of these considerations, this paragraph should be adapted to take account of the relative approach to the concept of personal data (suggested modifications **in bold**):

---

enable it to identify the data subject with additional data which the internet service provider has about that person.

[2] 48. In those circumstances, the VIN constitutes personal data, within the meaning of Article 4(1) of the GDPR, of the natural person referred to in that certificate, in so far as the person who has access to it may have means enabling him to use it to identify the owner of the vehicle to which it relates or the person who may use that vehicle on a legal basis other than that of owner.

[3] 100. Therefore, pursuant to paragraph 44 of the judgment of 19 October 2016, *Breyer* (C-582/14, EU:C:2016:779), cited in paragraph 91 above, it was for the EDPS to examine whether the comments transmitted to Deloitte constituted personal data for Deloitte.

104. It is apparent from paragraph 45 of the judgment of 19 October 2016, *Breyer* (C-582/14, EU:C:2016:779), cited in paragraph 92 above, that it was for the EDPS to determine whether the possibility of combining the information that had been transmitted to Deloitte with the additional information held by the SRB constituted a means likely reasonably to be used by Deloitte to identify the authors of the comments.

105. Therefore, since the EDPS did not investigate whether Deloitte had legal means available to it which could in practice enable it to access the additional information necessary to re-identify the authors of the comments, the EDPS could not conclude that the information transmitted to Deloitte constituted information relating to an 'identifiable natural person' within the meaning of Article 3(1) of Regulation 2018/1725.

22. *Pseudonymised data, which could be attributed to a natural person by the use of additional information, is to be considered information on an identifiable natural person, and is therefore personal,* ***according to recital 26 of the GDPR.***
This statement ***may also apply*** *if pseudonymised data and additional information are not in the hands of the same person,* ***insofar as the data subjects are identifiable****.* ***To qualify the nature of the pseudonymised data, it is necessary to put oneself in the relevant organisation's position in order to determine whether the information transmitted to it relates to 'identifiable persons.***
If pseudonymised data and additional information could be combined having regard to the ***actual and legal*** *means reasonably likely to be used by the controller or by another person, then the pseudonymised data is personal.* ***However, if the risk of identification is non-existent or insignificant, the pseudonymised data can escape the classification of personal data, notwithstanding the fact that the additional identification data have not been completely erased.***

Removing the additional information can make the data anonymous, just as not having and not being able to access the additional information in the first place can.

When Organization B (the recipient) accesses and processes data related to non-identifiable individuals, transmitted by Organization A, which holds identifiable information, the data processing:

- Falls within the scope of the GDPR for Organization A.
- Does not fall within the scope of the GDPR for Organization B.

This distinction influences the data governance approach that both organizations must adopt in compliance with applicable regulations.

2. Regulatory implications of a contextual approach to pseudonymised data

In light of the elements discussed in the previous section, the relative approach to the concept of personal data supports the conclusion that the use of data pseudonymization by a data controller (Organization A), and more specifically, the transmission of non-identifiable data to service providers (Organizations B), ensures compliance with the fundamental principles of the GDPR from the sender's perspective, without the transfers or the recipient's processing falling within the scope of the GDPR.

In this regard, Doctrine adheres to the legal reasoning set forth by Advocate General Spielmann in Opinion on Case C-413/23 P :

> 58. [...] The fact that the rules stemming from Regulation 2018/1725 do not apply to data relating to non-identifiable persons would not preclude entities that are at the origin of misconduct from incurring legal liability where appropriate, for example in the event of disclosure of data resulting in harm. On the other hand, **it seems to me disproportionate to impose on an entity, which could not reasonably identify the data subjects, obligations arising from Regulation 2018/1725, obligations which that entity could not, in theory, comply with or which would specifically require it to attempt to identify the data subjects**.

In practice, the relative approach to the concept of personal data has a significant impact on both the sender and the recipient of pseudonymized data, provided that the transfer is limited to non-identifiable data.

a. From the sender's point of view

For the data controller, acting as the sender of the data and retaining the additional information, compliance with GDPR obligations remains essential, as pseudonymized data continues to be considered personal data under the regulation.

The implementation of pseudonymization and the transfer of non-identifiable data to processors ensure compliance with the principle of data minimization, enhance security by mitigating risks of data breaches and unauthorized processing, and serve as an essential safeguard for data transfers, particularly to third countries outside the EEA, in line with GDPR requirements.

To provide a few illustrative examples:

- A SaaS platform, acting as a data controller, such as Doctrine, may engage a hosting provider, like AWS, as a data processor for the storage of its clients' data. In order to comply with core GDPR principles, the data controller can implement measures that ensure the hosting provider processes only encrypted data, with no access to the encryption keys, which remain under the exclusive control of the data controller. In this context, the risks associated with the data, including those concerning the rights and freedoms of data subjects, are significantly minimized—or even eliminated—through the use of pseudonymization.

- Similarly, a legal intelligence platform like Doctrine offers its users generative AI functionalities. These rely on Microsoft's Azure Europe GPT service, which incorporates Retrieval-Augmented Generation (RAG), while ensuring that no directly identifiable user data is transferred. Furthermore, legal queries are reformulated prior to processing. Consequently, no personal data is shared with the generative AI provider. As noted above, the risks to the rights and freedoms of data subjects are substantially reduced—or entirely mitigated—through the implementation of pseudonymization.

In both cases, the data controller performs pseudonymization of the personal data it collects for the provision of its services.

Pseudonymization provides a safeguard for personal data protection while supporting innovation. **Therefore, it is essential to establish clear guidelines regarding the impact of transferring pseudonymized data, both for the data subjects and for the stakeholders in the data processing chain**.

b. From the recipient's point of view

In both scenarios previously provided, there is no doubt that one could assert that the service providers are not recipients of any personal data. Thus, the risk of access by third parties, whether maliciously or through a request from a foreign government, as well as the risk of purpose diversion by the third party, such as reuse for training purposes, is minimized or even eliminated.

Beyond these considerations, the service provider cannot be held responsible for the obligations set out by the GDPR.