European Law Blog

The EDPB 01/2025 Guidelines on Pseudonymisation: A Step in the Right Direction?

Sophie Stalla-Bourdillon

European Law Blog

Published on: Feb 04, 2025

DOI: https://doi.org/10.21428/9885764c.3d9978a3

License: Creative Commons Attribution-ShareAlike 4.0 International License (CC-BY-SA 4.0)

The European Data Protection Board Guidelines on pseudonymisation were adopted on January 16, 2025, and are open for consultation until February 28, 2025. The EDPB has been working on this piece of guidance for several years. Initially, they were intended to be part of a broader set of guidelines dedicated to both anonymisation and pseudonymisation, in the wake of what had been done prior to the adoption of the General Data Protection Regulation (GDPR) with Opinion 05/2014 on anonymisation techniques. But it was then decided that two sets of guidelines should be adopted in the EDPB work programme adopted on 14 February 2023: one dedicated to pseudonymisation and one dedicated to anonymisation (You can compare with the 2021/2022 work programme).

In its Second Report on the application of the GDPR, the European Commission had stressed "the need for additional guidelines, in particular on anonymisation and pseudonymisation" raised by stakeholders and the Council, despite the existence of rich guidance produced by ENISA (e.g., in 2019, 2021, and 2022), national supervisory authorities (e.g., CNIL, AEPD together with the EDPS here and here, the DPC), as well as attempts to synthesise approaches at the global level. A recent document produced at the G7 level tries to highlight the differences in approach among G7 countries.

Given the depth of existing guidance on pseudonymisation, the added value of the EDPB Guidelines was expected to lie in their interpretation of the legal test for pseudonymisation, of which definition is found in Article 4(5) GDPR, to be read together with Recitals 26 and 28 GDPR. This clarification was particularly eagerly awaited in light of the position taken by the regulator of a key sector—namely, the European Medicines Agency in its external guidance on the implementation of the European Medicines Agency policy on the publication of clinical data for medicinal products for human use.

The case law on the notion of identifiability has evolved progressively. At least three cases would seem at first glance worth considering: *Breyer*, *Scania*, and *IAB Europe*. What is particularly striking, however, is that there is no mention of these cases in Guidelines 01/2025. Yet, the first part of the Guidelines is clearly an attempt to interpret the concept of pseudonymisation within the meaning of the GDPR.

The purpose of this blog post is twofold. First, it aims to highlight a few internal inconsistencies that make the Guidelines challenging to comprehend for those seeking a consistent approach to identifiability as both a technical and legal concept. Second, it seeks to shed light on the evolution of the legal test for identifiability, to highlight convergences between the approaches of the EDPB and that of the CJEU. But before highlighting internal inconsistencies and convergences with the CJEU's approach, let's summarise the approach taken by the EDPB in its Guidelines.

The EDPB's Approach

Unsurprisingly, the EDPB presents pseudonymisation as a key **compliance measure**. The EDPB makes it clear that, although pseudonymisation is usually implemented to meet the data protection goal of confidentiality, it can also serve other data protection goals such as accuracy, purpose limitation and fairness.

Just like anonymisation, pseudonymisation is fundamentally a trade-off between different types of data protection goals, as acknowledged implicitly by the EDPB: while pseudonymisation may help preserve confidentiality, it can also facilitate linkability (e.g., the linkability of records included in different data sources but pertaining to the same individual) and reuse. If the party reusing the pseudonymised data does not have access to the identifying additional information, data subjects may have less opportunities to intervene and exercise their rights, as Article 11(2) GDPR provides that if the controller is considered not to be in a position to identify the data subject, Articles 15 to 20 shall not apply except where the data subject provides additional information enabling his or her identification. Of note, as Article 11(2) GDPR does not foresee that the right to object may not apply, it would be useful to include some examples illustrating how it can be preserved.

The strength of pseudonymisation when used as a compliance measure to reduce re-identification risks is determined by introducing the concept of **pseudonymisation domain**.

Back in 2017, the concept of domain had been introduced in an attempt to conceptualise differences between anonymisation, Article 11 de-identification and pseudonymisation.

A pseudonymisation domain is defined by the EDPB as an "environment in which the controller or processor wishes to preclude [the] attribution of data to specific data subjects." The range of situationally relevant entities against which confidentiality risks are being mitigated through the data transformation process is thus defined by the party responsible for performing the pseudonymisation. This range is a key consideration for evaluating the robustness of the pseudonymisation process. When external attackers, or more widely unanticipated recipients of the data, are not taken into account, the level of residual re-identification risks is likely to be higher than if these entities are taken into account to select such controls.

Importantly, by definition, a pseudonymisation domain does not include the entities "authorised to process additional data allowing the attribution of the pseudonymised data to data subjects." (see definition p. 46)

Why exclude the entities authorised to access the additional information from the pseudonymisation domain? This is because including them would mean performing an anonymisation process as opposed to a pseudonymisation process, as hinted in Recital 26 GDPR. The EDPB's task in these Guidelines is not to explore situations in which a pseudonymisation process could lead to an anonymisation process, although it clearly and logically states "[e]ven if all additional information retained by the pseudonymising controller has been erased, the pseudonymised data can be considered anonymous only if the conditions for anonymity are met." (para. 22) Note that the EDPB does not write that anonymisation is only achieved if the additional information retained by the pseudonymising controller is erased.

The EDPB Guidelines include in its annex several detailed examples of the application of pseudonymisation. For each example the EDPB adopts a systematic approach, which essentially illustrates how individual-level data can be transformed into pseudonymised data. It is not expressly stated in the Guidelines that pseudonymised data can never lead to anonymised data. In other words, there is no express rejection of a risk-based approach to anonymisation. A thorough evaluation of the risks should remain an option, e.g. possibly when singling out risks persists within the data, as recalled by the EDPB itself in Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models.

Finally, the paragraphs dedicated to international data transfers echo the description of Use Case 2 found in the recommendations on supplementary measures issued in 2021. The EDPB confirms that a whole gamut of data transformation techniques will have to be applied, which are likely to include generalising techniques such as k-anonymisation (a privacy-preserving method ensuring that each individual in a dataset is indistinguishable from at least k−1 other individuals with respect to certain attributes deemed to be quasi-identifiers).

Internal Inconsistencies

The first conceptual challenge emerging from the Guidelines comes from the confusion that is made between the ‘identifiability’ and the ‘relating to’ criteria. As a reminder, the EDPB’s predecessor had broken down the Data Protection Directive’s definition of personal data into a four-prong test in its 2007 Opinion on Personal Data: ‘any information’, ‘relating to’, ‘identified or identifiable’, ‘natural person.’ Pseudonymisation as a data transformation technique that aims to pursue (at least in part) the data protection goal of confidentiality, has no implication for the appreciation of the ‘relate to’ criterion which tries to answer the question whether the data describes an individual or something else, like an event, or a machine, or say an animal. As such, pseudonymisation only impacts the identifiability level associated with the individual record. It is therefore confusing to write that “to attribute data to a specific (identified) person means to establish that the data relate to that person.” (para. 17)

A more precise formulation would be: '[a]ttributing data to a specific individual means determining that the individual is either identified or identifiable based on the data available within the pseudonymisation domain.”

The technical literature on de-identification usually draws a distinction between direct and quasi (or indirect) identifiers to explain the difference between direct and indirect identifiability, which seems to be at the heart of Article 4(1) GDPR. Although the EDPB draws a distinction between direct and quasi-identifiers, the terminology could appear confusing. The EDPB defines direct identifiers as essentially distinguishing references. However, a pseudonym is by definition a unique reference and therefore distinguishing. This would mean that pseudonymised data is always directly identifying, which is not exactly what that EDPB is trying to say about pseudonymised data. At para. 8, the EDPB write that “it is clear that direct identifiers need to be removed from data if those data are not to be attributed to individuals.” To make sense of what the EDPB is

saying, one would need to add that certain types of direct identifiers are not identifying, which is a confusing assertion.

A better formulation would therefore imply acknowledging that direct identifiers have two key characteristics: distinguishability (i.e., uniqueness) and availability (they are potentially available or accessible to or by an attacker). When appropriate data segmentation measures have been implemented, and considering the pseudonymisation domain only, pseudonyms should not be considered available.

For the sake of clarity, it may help to include two sets of definitions: one for direct identifiers and one for quasi or indirect identifiers.

One last point on international data transfers. As explained here, the description of what pseudonymisation processes should look like in the context of international data transfers seems to suggest that no thorough evaluation of the risks is ever possible in this context. In addition, it is not clear whether the EDPB assumes that third-party public authorities should be considered as having some form of prior knowledge or not. If that's the case a more detailed explanation as to why this is the case would be useful, as mentioned here.

Compatibility with CJEU Case Law

Many have criticised the EDPB Guidelines stating that it relies upon a misconception of the legal test for identifiability.

While it is true that the EDPB does not perform an analysis of the CJEU case law, the EDPB's approach and that of the CJEU as it stands today do not seem to be misaligned. Truly, the CJEU is still in the process of refining the identifiability test under the GDPR, as an appeal judgment on this matter is still expected in the SRB case. Looking at the CJEU case law on identifiability though, there seems to be a way to make sense of both the CJEU case law and the EDPB approach to pseudonymisation and arguably anonymisation as well. This can be done by referring to the concepts of distinguishability and availability introduced earlier. Let's explain. For a detailed overview of the CJEU case law through the lenses of these two concepts see my recent paper available here.

In Opinion 01/2025, the EDPB is essentially saying (assuming it manages to streamline its definitions) that within the pseudonymisation domain, pseudonyms are distinguishing but not available. As a matter of principle, this does not exclude that if a thorough evaluation of the risks is conducted, transformed data within an anticipated recipient's controlled environment could never be considered anonymised. However, until such a demonstration is made—bearing in mind that the burden of proof lies with the party claiming the anonymised status—the data should be regarded as pseudonymised. What is more, feedback loops, i.e., whether it is anticipated that the pseudonymised data will enrich the original data at some point in time, are also relevant for the analysis. Each time a feedback loop is maintained between the original data and the pseudonymised data,

there are good reasons to adopt a holistic approach for the legal assessment and not to artificially separate the pseudonymising entity's hands from the data recipient's hands.

Of note, in *SRB* there is no demonstration that a thorough analysis of risks has been performed and there is a feedback loop that is maintained between the original data and the transformed data.

In *Breyer* and *Scania*, the CJEU considers the status of two types of data points: dynamic IP addresses and Vehicle Identification Numbers (VINs). Importantly, IP addresses and VINs are not pseudonyms as the EDPB views them. What is more, in *Breyer*, dynamic IP addresses are considered to be both distinguishable (singling out takes place) and available (the data holder, i.e., an online service provider, has the legal means to access additional identifying information). In *Scania*, VINs are considered indirect personal data in the hands of vehicle manufacturers, which is essentially implying that they are distinguishing and potentially available to anticipated recipients, i.e., independent operators.

In *IAB Europe*, the CJEU adds a very important nuance, which suggests that the concept of personal data is rightly **functional**, as stated here. When the anticipated processing implies or enables the profiling of data subjects, the only criterion that matters is distinguishability. What this implies is that a thorough evaluation of the risks is in this case irrelevant, which aligns with a high level of data protection.

And in Bindl, although the General Court's reasoning lacks nuances, the description of the last disputed data transfer seems to imply that the IP address at stake is both distinguishable and available.

To conclude, while a few key definitions still need to be refined, Guidelines 01/2025 represent a step in the right direction. Their significance is set to grow with the imminent entry into force of the European Health Data Space Regulation, as Article 66 EHSD provides that access to electronic health data for secondary use will happen either in an anonymised format or in a pseudonymised format. One final point that would require clarification is the methodology for conducting a thorough risk evaluation, which will be fiercely debated, as illustrated by the exchanges having taking place in the Thin Database case decided by the Italian DPA.

Sophie Stalla-Bourdillon is co-Director of the Privacy Hub. She is also a visiting professor at the University of Southampton Law School of law, where she held the chair in IT law and Data Governance until 2022. She was Principal Legal Engineer at Immuta Research for six years. Sophie is the author and co-author of several legal articles, chapters and books on data protection and privacy. She is Editor-in-chief of the Computer Law and Security Review, a leading international journal of technology law, and has also served as a legal and data privacy expert for the European Commission, the Council of Europe, the Organisation for the Cooperation and Security in Europe, and for the Organisation for Economic Cooperation and Development.