Legal Knowledge and Information Systems
R. Markovich et al. (Eds.)

doi:10.3233/FAIA251589

How Anonymous Is Anonymous? A Techno-Legal Exploration

Stephanie VON MALTZAN [a,1], Daniel VOGEL [b], Marc OHM [b,c] and Florian IDELBERGER [a]

[a] *Karlsruhe Institute of Technology and FIZ Karlsruhe*
[b] *University of Bonn and Lamarr Institute*
[c] *Fraunhofer FKIE*

ORCiD ID: Stephanie von Maltzan https://orcid.org/0009-0004-8135-4202, Daniel Vogel https://orcid.org/0009-0004-1717-1524, Marc Ohm https://orcid.org/0000-0002-2913-5270, Florian Idelberger https://orcid.org/0000-0002-9253-1524

Abstract. This paper proposes a pragmatic exploration to facilitate the categorisation of personal data as anonymous, quasi-anonymous, or pseudonymous, emphasising contextualised threat modelling and proportionality over binary thresholds. Using an integrated legal analysis and system-level threat model, we map legal criteria to the design features and assess whether a privacy-preserving system like DROPS can credibly achieve anonymisation under the GDPR. This allows us to evaluate the discrepancy between the technical realities of maximising anonymisation techniques and the requirements for anonymisation stipulated by the EU data protection law corpus. The distinguishing feature of this paper is its grounding of the legal analysis in the technical architecture, thereby bridging the gap between abstract regulation and system-level design. This demonstration has the potential to serve as a model for enhancing data protection measures, particularly for entities that handle high-risk or otherwise sensitive data and for regulators to issue new concrete guidance on anonymisation.

Keywords. Anonymisation, Quasi-Anonymisation, Pseudonymisation, Hashing, Privacy, GDPR, PII data

1. Introduction

Amid a rising tide of data breaches and increasingly sophisticated cyberattacks, determining whether personal data has been genuinely anonymised under the General Data Protection Regulation (GDPR) has become a complex and pressing issue. The GDPR does not provide an explicit definition of anonymous data. Instead, it can be inferred from a combined interpretation of Article 4(1) GDPR, which defines personal data, and Recital 26 GDPR, which explains how to assess identifiability. However, the threshold between anonymisation and pseudonymisation remains unclear and debated. Data protection authorities, courts, and legal scholars across Europe have adopted varying interpretations – from

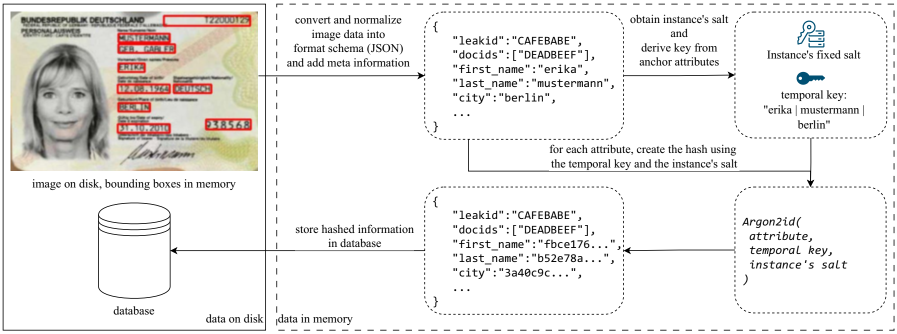
[1] Corresponding Author: Stephanie von Maltzan, stephanie.maltzan@kit.edu

relative to absolute approaches and from strict to functional anonymisation [21,35,26,10] —- and legal and technical experts often view anonymisation differently. In light of this, many organisations either incorrectly assess the effectiveness of their anonymisation processes or struggle to comply. Against this backdrop of uncertainty, this paper offers a techno-legal exploration of anonymisation and its role in data protection regulations. It emphasises the importance of contextual threat modelling and proportionality, rather than adopting an all-or-nothing approach based on binary thresholds such as the absolute and relative approach. Using the DROPS system [14] as a case study, we ask: can a system that employs cutting-edge, privacy-preserving methods credibly claim to anonymise data under the GDPR, or does the data remain personal despite strong safeguards? This is of particular interest given the different views [18,11] on whether pseudonymisation should be considered an effective means of anonymisation. If strict anonymisation cannot be practically achieved, even with robust safeguards in place, should the legal standard be adjusted to take a more pragmatic, risk-based approach? Our analysis underpins legal discussions by examining the technical architecture of DROPS, thus linking abstract regulatory standards to system-level design choices. This integrated approach clarifies GDPR compliance and serves as a model for data protection authorities to issue new, more concrete guidance, as well as enabling the European Court of Justice (CJEU) to consider illegal means in its assessments. We argue that absolute and relative binary categories no longer provide an adequate framework for addressing identifiability under the GDPR. Although we agree with the recent SRB judgement [11] that pseudonymous data can be anonymous in certain instances, we also recognise that illegal means to obtain additional information should be considered. Ultimately, our aim is to foster a more practical understanding of anonymisation through this example.

The paper proceeds as follows: Section 2 reviews work relating to the intersection of anonymisation in law and technology. Section 3 describes the DROPS system architecture and design choices, while the subsequent section assesses DROPS from legal and technical perspectives, applying GDPR and WP29 anonymisation criteria in light of a threat model.

2. Related Work

Prior research [28,38,27,39,16,12,33,6,34] has addressed specific aspects of the problem, such as legal critiques of anonymisation under the GDPR [4,29,32,37,31], technical limitations of anonymisation techniques [6,15,22,24], and highlighted key misunderstandings[2]. Achieving meaningful anonymisation in practice remains a significant challenge [25,36,37] that straddles law and computer science. The most influential regulatory guidance remains the Article 29 Working Party's (WP29) Opinion 05/2014 [3], which introduced a three-part test for anonymisation in conjunction with the "'means" test.[36] In its latest opinions (Opinion 28/2024 [17] and Opinion 01/2025 [18]), the European Data Protection Board (EDPB) has further clarified the "'means" test. Case law of the CJEU – notably the Breyer [7], IAB [10], Scania [8], Olaf [9] and SRB [11] decisions, — has added nuance (and also uncertainty) by interpreting identifiability based on access to additional information, as well as considering whether re-identification by third parties is legally feasible[40]. The interpretative nature of the provision, combined with inconsistent guidelines and case law, has not only led to divergent views among legal scholars but also among regulators. For example, Spain's [1], France's [5] and Italy's [23] data protection authorities tend



```
{
  "leakid":"CAFEBABE",
  "docids":["DEADBEEF"],
  "first_name":"erika",
  "last_name":"mustermann",
  "city":"berlin",
  ...
}
```

```
Argon2id(
 attribute,
 temporal key,
 instance's salt
)
```

```
{
  "leakid":"CAFEBABE",
  "docids":["DEADBEEF"],
  "first_name":"fbce176...",
  "last_name":"b52e78a...",
  "city":"3a40c9c...",
  ...
}
```

BUNDESREPUBLIK DEUTSCHLAND
PERSONALAUSWEIS
T22000129
MUSTERMANN
GEB. GABLER
ERIKA
12.08.1964
DEUTSCH
BERLIN
31.10.2010
238568
image on disk, bounding boxes in memory
database
data on disk
convert and normalize image data into format schema (JSON) and add meta information
obtain instance's salt and derive key from anchor attributes
Instance's fixed salt
temporal key: "erika | mustermann | berlin"
for each attribute, create the hash using the temporal key and the instance's salt
store hashed information in database
data in memory

Figure 1.: Flow diagram of data processing within the system.

toward a strict standard, whereas Ireland's [13] has been more lenient in some instances. Such fragmentation underscores the need for a harmonised approach grounded in both sound legal reasoning and practical feasibility, as proposed in this paper. To date, existing research has relied solely on regulatory guidelines, case law or a technical approach, generally accepting the unresolved status quo without conducting an in-depth techno-legal analysis based on a concrete threat model. A key challenge is to demonstrate at what point pseudonymised data might be considered effectively anonymised, and where its ragged edges lie. This paper builds on existing literature by synthesising the perspectives of law and technology and seeks to fill the gap by evaluating the DROPS system against legal criteria using a thread-scenario-based assessment [20], thereby illustrating how and when the GDPR's anonymisation threshold might be reached (or missed) in practice.

3. Architecture

The DROPS system[14] is designed to enable comparisons of an individual's data with records of known data breaches (leaks) without exposing any party's raw data to the other. The system accomplishes this through a pipeline of cryptographic hashing and secure comparison protocols, described below and illustrated in Figure 1.

Each personal data value observable in raw identity data is separately hashed using a combination of a system-wide randomly generated secret salt and an anchor, creating unique hashes that do not inherently reveal their link to an anchor to outsiders. The anchor is a value derived from a fixed set of attributes (three in total) for each individual (for instance, a combination of name and address), allowing structured pseudonymisation serving as a identity-unique temporal key. All personal data from a leak dataset is hashed using the memory-hard Argon2id algorithm with the system's secret salt, as recommended by OWASP [30], then truncated. Argon2id's CPU- and memory-intensive nature adds a significant barrier to brute-force or dictionary attacks on the hashes. Importantly, after the hashes are computed, the leak dataset, the original plaintext data (preimage) and the anchor values are immediately discarded. The DROPS database stores only the salted hashes of attributes, each tagged with an internal document ID (doc_id) and a leak ID indicating from which leak the record stems. Neither the anchor nor the preimage is stored in the database, and neither is needed for it to operate. Truncation limits the information an attacker can gather from any single hash value, limiting the effect of precomputed tables, while only posing negligible collision risk. In fact, truncated Argon2id hashes still enable almost 100 percent accurate matching between datasets through cryptographic protocols like Private Set Intersection (PSI), allowing DROPS and its users to achieve near-perfect intersection accuracy.

DROPS employs a form of PSI to enable secure matching of records between a querying party and the leak database without revealing sensitive data to either side. To initiate a query, an authorised user must use the DROPS client app, which handles the following process client side given correct input: Give anchor attribute values and an ordered list of identity attributes and these attribute's values. Hash each value list entry using anchor as key and the system's salt. Hash the concatenation of these hashes again using the system's salt, then truncate the resulting hash to a pre-negotiated length. Query with truncated hash and ordered list of identity attributes. Upon receiving the query, DROPS compares truncated hashes of attribute sets that share doc_ids. A match occurs only if the full set of queried attribute hashes exists within the DROPS database, assuming no collisions. Throughout this PSI-based protocol, no raw personal data is ever exchanged. The querying party only learns whether the queried attribute hashes exist within the DROPS database. Conversely, the DROPS server learns nothing about the query other than the fact that a query was made – it cannot derive which person the query pertains to.

Through these measures, DROPS establishes a highly secure and privacy-preserving environment. It stores only pseudonymised personal data, requires prior knowledge of key attributes to generate queries, and employs cryptographic protocols so that information disclosure is minimal and controlled. Subsequently, we evaluate how these technical choices translate to GDPR compliance in terms of identifiability of individuals in the data.

4. Techno-legal assessment

Under the GDPR, data are considered anonymous if individuals are ”not or no longer identifiable” taking into account ”all the means reasonably likely to be used” for re-identification. WP29's Opinion [3] established a standard three-criterion test: after anonymisation, it should be impossible to (1) single out an individual in the dataset, (2) link records relating to the same individual (within the dataset or between datasets) relating to the same individual, and (3) infer any additional information about an individual from the data. If any of these risks remain, the dataset cannot be regarded as fully anonymised. Instead, the controller must evaluate the residual risk of identification and determine if it is acceptably low in practice [3,17,20]. Importantly, the GDPR does not demand zero risk — an impossible standard — but any risk of re-identification must be extremely remote. Data can be considered effectively anonymised if re-identifying a person would require a disproportionate effort. This assessment requires consideration of all the means that could reasonably likely to be used by the controller or a third party to identify individuals, and the determination of those means should be based on objective factors such as the cost of identification, time required and technologies available at the time of processing. Anonymisation procedures and their evolution must therefore be continuously reviewed and evaluated [2]. This test requires a comprehensive and contextual risk assessment that goes beyond simply assessing the technical possibility of re-identification and takes into account the practical feasibility and likelihood of such attempts. Consequently, this raises the question of whether there exists additional information, either currently available or that may become available in the future, that could be utilised in conjunction with the existing data to identify the individual in question effectively. Controllers should first focus on the concrete means (here: original leak dataset, anchor, salt) that would be required to reverse the anonymisation technique, particularly concerning the costs and know-how required to implement these means and the assessment of their likelihood and severity. DROPS must, thus, balance its anonymisation efforts and costs (both in terms of time and resources required) against, for example, the increasingly low-cost availability of technical means to identify individuals in datasets, the public availability of other datasets and the leakage of secret keys, as well as the costs of brute-force attacks. EDPB stresses that risk assessment should account for all actors, including malicious attempts – not only the "legitimate" data recipients foreseen by the controller [18]. This broad view contrasts with the CJEU interpretations [7,8,10,19], which to date have not been particularly explicit about the scope of relevant other persons beyond data recipients concerning the application of the means test. Essentially, courts focus on the intended or anticipated data flows and apparently consider only legal means as applicable. However, this standpoint may be excessively restrictive insofar as it poses a considerable real threat to privacy, especially in a world of frequent data breaches. Following the EDBP's [18] approach, when determining whether data is truly anonymised, it is necessary to consider not only honest actors but also individuals who might deliberately attempt to circumvent de-identification. Our assessment of DROPS follows this approach, examining identifiability from multiple threat angles.

4.1. Singling out

Singling out is "the ability to locate or isolate an individual's record in the dataset" [10]. Even without a person's name, if one can pinpoint a unique record (or set of attributes) that corresponds to the same individual, that individual is singled out.[3] **Access to the DROPS database:** All personal records are individually stored as Argon2id anchor-based hashes along a corresponding `doc_id` (and `leak_id`) to group all attribute-hashes from the same document source. Each hashed identity record corresponds to one real person's attribute (e.g. name) from the leaks. Because DROPS uses deterministic hashing with a fixed salt and anchor per person, each individual's attribute data yields a unique pseudonym (per `doc_id` a set of pseudonymised values) that acts as an identifier for that person within the system. This means that, even though the actual identifiers are hashed, an attacker with access to the DROPS database can use them to distinguish one person's record(s) from another's because each record is stored alongside its `doc_id`. Attackers do not need to know the person's name or the details behind the records to distinguish them. WP29 explicitly noted this pitfall [3] The structure of DROPS inherently allows for singling out individual records in the dataset.

Through authorised use (a client with query access): If the user's query finds a match in DROPS, the user has effectively singled out that individual as present in the leak. This is an intended feature outcome, not a bug – but it does highlight that an individual can be singled out (identified as a breach victim) by combining the DROPS processing with the additional knowledge held by the contracting organisation (the individual's identity). From an identifiability perspective, the system deliberately allows known individuals to be singled out (alerted), albeit only by someone who already knew their identity. Therefore, DROPS does not eliminate the risk of singling out, but rather limits who can singling out records. Notably, the authorised user already knows the identity, DROPS did not reveal a new identity, but it did confirm the existence of that person's data. From a strict legal perspective (although the ability to do anything useful with a singled-out record is severely constrained), DROPS does not eliminate the risk of singling out.

4.2. Linkability

Linkability refers to the ability to link two or more records concerning the same individual within the same dataset or across different datasets [3]. For data to be anonymised, it should not be possible to correlate separate data points and conclude that they are about the same data subject.

Access to the DROPS database: The DROPS system's functionality relies on linkability by design. The anchor-based hashing scheme links all attributes of a record under a common key (the anchor, although it is not stored). Internally, this is reflected in the `doc_id` grouping in the database. This internal linking is necessary for the system to function – it allows the PSI query to check for a matching set of attributes. If the same individual's information appears in multiple leaks imported into DROPS, the system will, as long as the personal and anchor information is consistent, assign them the same pseudonym (hash) each time. These hashes will then be assigned multiple `leak_ids`. An insider who inspects the database could notice that two different `doc_id` groups have identical hash values, suggesting that they belong to the same person. This means that DROPS itself can inherently link records across different leaks that pertain to the same individual. Even though the person's name is not stored, the pattern of recurrence is visible. This internal linkability is necessary for the system's purpose. However, this means the second WP29 criterion is not met – the system intentionally preserves the ability to link records referring to the same individual.

Through authorised use (a client with query access): Linkability is the core functionality provided to the querying party. When an organisation queries "Jane Doe, Main Street, jd@example.com" (hashed) and DROPS finds a match, it is linking that query record to a leak record. The PSI protocol outputs the intersection, i.e., the links between the set of query hashes and the set of leak hashes. In this way, the authorised user learns that the record in their database (Jane Doe) corresponds to a record in the DROPS database. This linking is done in a privacy-preserving manner (neither side learns anything beyond the existence of the link), but it is linkable across datasets nonetheless (the querying party's list and the DROPS database). A malicious authorised user might try to exploit this by, for example, querying many different partial identities to see which ones exist in DROPS (a fishing expedition to find out who is in the database). DROPS mitigates such abuse by requiring the full anchor for queries and by monitoring for suspicious query patterns. This makes it difficult to systematically probe for links without already knowing the identity of a target individual.

External attacker: What complicates linkage for outsiders is that the anchor is secret, and the salt is system-wide but not public. If an attacker did not know about the `doc_id` link, they might not realise that two different-looking hashes actually belong to the same person. However, DROPS uses pseudonyms for different attributes of the same person within the system using the same anchor – for example, all of Jane Doe's data (e.g. name, address) becomes hashed identities under the same anchor context. Even truncated hashes won't break this link, as the hash input data are still linked through the anchor. As the system maintains consistent pseudonyms the second criterion is not met. This is in line with WP29, which noted that when the same key is used for an individual across records, linkage is "trivial" [3]. An attacker could link records that belong to the same person either by querying using corresponding known personal data and finding a match or by getting access to the DROPS database and filtering by `doc_id`. Neither of these methods, however, allows the attacker to learn any identifiable information that they doesn't already know.

4.3. *Inference*

Inference is the ability to deduce the value of an attribute with significant probability using other information [3]. DROPS severely limits what can be inferred by anyone without prior knowledge. The data stored and exchanged is heavily transformed (hashed), so that personal attributes cannot be directly inferred; the hash reveals nothing semantic other than its appointed attribute. The data flows are, furthermore, sufficiently controlled that unauthorised inference by outsiders is impossible (since they see nothing but hashed tokens). The authorised parties (the clients), however, infer something non-trivial: they learn that a particular user's credentials were found in a leak. This is arguably new information about that user. If a malicious client identifies the existence of a person's data in a leak, that client could try to find the leak. In fact, the existence of other entries could even be verified by using the leak to gather information about other persons and query their presence in the DROPS database. In that case, the additional knowledge of the fact that the rest of the leaked content is also part of the DROPS database is likely to be insignificant. DROPS performs well against the inference criterion against outsiders – it is extremely resistant to someone inferring new information about individuals. However, under WP29's strict test, the existence of any plausible inference attack means that the data is not anonymised.

By the strict criteria, DROPS does not fully protect against singling out, linkage and inference, and thus cannot be considered anonymisation by this standard. The CJEU's stance that only legal means must be employed would not alter this.

4.4. *The "all means reasonably likely to be used" test*

While the analysis above finds that DROPS does not satisfy the WP29's strict criteria, the WP216 opinion also states that controllers should carry out an evaluation of the identification risks in order to determine whether the residual risk is acceptable; in other words, whether the anonymisation process is sufficiently robust [3]. For the purposes of this paper, the following simplified analysis considers several attacker models and evaluates the effort, means and likelihood of success for each. This analysis brings us to the concept of quasi-anonymisation, which we introduce to describe data that is not strictly anonymous in law, but is sufficiently de-identified under a contextual, proportionate risk analysis.

External Attacker: No Access to Salt or Anchor Given that an attacker gains access to the DROPS database and has access to all hashes associated with their `doc_id` and `leak_id`. An external attacker with sole access to the DROPS database has no direct means of re-identifying individuals using the DROPS database alone. If the attacker had access to various leaked data, some of which were used for the DROPS database, the attacker could try to create hashes based on the known leaked data. These hashes could be used to find matching hashes in the database, thereby de-pseudonymising the entry. Without knowledge of the salt, re-identification would require the attacker to guess the salt, which would feasibly require an exhaustive search through all possible salts. Since Argon2id is set up to limit the amount of hashes per second that can be calculated, finding the correct salt is a disproportionate effort, as shown by the following calculation: Assuming a DROPS system uses a salt with a length of 32 Byte or 256 Bit and that an individual record can be hashed with Argon2id in 0.01 seconds, meaning 100 hashes per second. An exhaustive search for all values of the salt would require an average of 4.4e68 years using one processor. As the database stores only truncated hashes, the attacker would also have to verify a matching salt for one record with at least a second record, adding an additional step that is somewhat negligible considering the hashing cost. Given the computational intensity of Argon2id and the entropy added by both the salt and the anchor, brute-force reversal is not practically feasible. This would require expensive computational resources. Recital 26 GDPR explicitly directs consideration of the cost and time required, taking into account current technology. Here, an external attacker would need astronomical resources to guess each unknown salt/anchor and recompute Argon2id hashes. Under the Recital 26 test, the means of such an attacker are not reasonably likely – the effort and cost of re-identifying identities are disproportionate, and the identification thus fails the threshold of objective identifiability. In short, to an external attacker without access to anchors and salt (additional information), the DROPS data – although technically pseudonymised — can be deemed, for all practical purposes, anonymous under current technological constraints: without the secret keys, the data cannot be linked to individuals by any feasible method. In terms of considering pseudonymous data as anonymous, this is consistent with the recent SRB judgment [11], but inconsistent with EDPB guidance [18]. We argue that the remaining risk appears to be acceptable. Recital 26, however, demands a holistic assessment, meaning that DROPS must also consider actors who might obtain the additional information.

Malicious Insider (Access to database, salt): If an attacker has access to both the DROPS database and the salt, either as a malicious insider or through a salt leak, the means of identification are much more accessible. This attacker would be enabled to generate hashes for identity records for which the anchor attributes are known. It would be possible to use acquired identity data to look for matching hashes, thereby inferring which identity records are stored within the DROPS database. Notably, the attacker does not learn any identity data that is not already known. Nevertheless, it is possible to attribute the data to a specific individual by leveraging additional information, such as a secret salt and the individual's identity details. From a legal perspective, the presence of an individual (the insider) who possesses or obtained a key to unlock the pseudonymised data indicates that the data cannot be regarded as fully anonymised, unless the effort expended is disproportionate. This illustrates that DROPS data is only as anonymous as its secret salt remains secret, as long as the salt and anchor knowledge are confined to the controller's internal system. However, if an insider reveals that secret, the hashes become identifiers, which necessitates treating such data as pseudonymised, with robust technical and organisational safeguards in place, though not to the extent of irreversible anonymisation. In summary, DROPS has been designed to make re-identification of the hash records a complex, but not disproportionately impossible, undertaking in the event of salt leakage. Beyond known identities, guessing unknown identity records involves the complexity of matching attributes and the correct anchor, which may be subject to hash collisions, taking hash truncation into account. It is arguably less disproportionate to find leaks presented to DROPS that researchers have found on the internet than to guess identity records. However, this would be impossible for leaks shared only by a whistleblower with DROPS and no third party. Without additional information, the complexity of making valid guesses increases significantly due to anchor-hash binding and hash truncation and would require disproportionate brute-force attacks or contextual guessing. Ultimately, without access to the original leak or external auxiliary data, an attacker cannot meaningfully narrow the input space. From a legal perspective, this constitutes a disproportionate effort, rendering the data effectively anonymous. However, if it could be possible to access the original leak, the threshold of disproportionate effort may not be reached. In such cases, the concept of quasi-anonymisation is particularly relevant and is used to describe data that is not strictly anonymous in a legal sense, but is considered sufficiently de-identified under a contextual and proportional risk assessment – especially when re-identification is close to the threshold of disproportionate effort. Even without full anonymisation and with a hashing technique that essentially amounts to pseudonymisation, the contextual risk of re-identification here is deemed low enough to nearly reach the threshold of disproportionate effort. As a result, the data can be classified as outside the scope of the GDPR's definition of personal data. Nevertheless, this status must undergo regular reassessment to ensure its ongoing relevance amid technological advancements and shifting threat landscapes, as well as to incorporate new information. Argon2id is currently the most robust option; however, this could change in the future. Anonymisation should not be seen as a one-off technical achievement but rather as a dynamic, context-sensitive process. Taking the CJEU's perspective into account, it would also be anonymous, since illegal means are not considered. However, according to the EDPB, pseudonymous techniques cannot be attributed to anonymous data.

Malicious Insider (Anchor, Salt, API key): Having access to a correct API key would allow an attacker to impersonate a DROPS client and create queries to attack the DROPS system either through database probing, DDoS or similar attacks, which we won't discuss as they are outside the scope of this paper. Probing the DROPS database using a known API key, salt, and prior knowledge of identity anchor information could be used to attempt to learn about the existence of specific attribute records of previously unknown identities. The DROPS database could be used as an oracle to verify a guess, provided the attribute is present in the database. When combined with the salt and anchor knowledge, this creates a feasible attack surface, especially when guessing small attribute spaces. Notwithstanding the implementation of technical and organisational safeguards, re-identification remains a realistic possibility. Consequently, the data remains pseudonymised personal data, albeit not from the perspective of the CJEU.

Authorised User of DROPS (Querying via Client): A malicious authorised user who interacts with the system only through the DROPS client would have access to an API key, as well as the DROPS client software. From a GDPR perspective, the limited and cryptographically shielded interaction reduces identifiability for the querying user to a negligible level. Users are constrained to queries for which they possess the anchor information, i.e., prior knowledge of the individual. Consequently, the users are unable to re-identify the individual, as they are already aware of their identity. As re-identification is predicated on pre-existing knowledge, and no further indirect identifiers or structural links are revealed, the DROPS data is (besides the querying hashes) functionally anonymous for the querying user, and vice versa for DROPS, while the DROPS data is pseudonymous for DROPS. This is in line with the IAB's [10] reasoning that the status of personal data is relative to the information held by each party: one entity's data may be considered personal, while another's may be anonymous.

Incorporating these technical scenarios into legal analysis reveals a pattern: Due to the intended use of DROPS, full anonymisation cannot (which is legally debatable under different approaches) be achieved in all cases – a situation that frequently arises when the data's utility is still necessary. DROPS occupies a middle ground: quasi-anonymous – the data is functionally anonymous given the realistic threat environment and existing safeguards, though not irrevocably anonymised against all attackers, reflecting a pragmatic and more practical balance between utility and privacy. Although residual risks persist, they are effectively mitigated.

Quasi-anonymisation acknowledges that, while full anonymisation may be unattainable in specific contexts, strong technical and organisational (even pseudonymisation techniques) safeguards can significantly reduce the risk of re-identification to an acceptable level. This distinction recognises that anonymisation is not a binary phenomenon, but rather a spectrum. It is proposed that, if quasi-anonymisation is achieved, the GDPR would not apply in advance, but rather steps would be taken to limit the compliance gap and associated risks of non-compliance with the transformation into personal data.

5. Conclusion

By applying the ”means” test through a simplified attacker cost-and-effort model, we demonstrated the different outcomes of applying the CJEU's and EDPB's guidance, and showed the importance of conducting context-specific risk analyses to determine when data can be considered effectively anonymised. The proposed concept of quasi-anonymisation as a legally relevant category highlights a pragmatic middle ground where data protection can be robust without requiring perfection. The effectiveness of anonymisation will be demonstrated further in future, quantitatively. This paper aims to encourage discussion on widespread practical anonymisation, encourage companies to adopt advanced data protection practices and regulators to reach a consensus. We argue that EDPB should provide concrete guidance that outlines, first, clear criteria for anonymisation and, second, the flexibility to implement quasi-anonymisation measures. If the criteria remains too absolute or unclear, there is a risk that controllers will either misapply it or disregard anonymisation efforts altogether. A more apparent acknowledgement of contextually‘”safe” data would incentivise the adoption of strong technical measures by offering some relief under the law when the re-identification risk is minimal. Future work and policy should focus on refining the concept of quasi-anonymisation and developing best practices for achieving it within the GDPR's framework. Importantly, our assessment also sounds a note of caution: methods used in DROPS, while significantly reducing risk, do not entirely eliminate the possibility of re-identification, especially if additional information becomes available. Controllers should avoid any false sense of security. Relying on a single technique is not enough; instead, a defence-in-depth strategy is imperative. Such a holistic approach is essential to maintain effective anonymisation over time, ensuring compliance with the GDPR's demands for data protection by design and by default, and preventing non-compliance in the event of the transformation of data into personal data.

References

[1] AEPD. Anonymization III: The risk of re-identification, 2025. Accessed August 10, 2025. https://www.aepd.es/en/prensa-y-comunicacion/blog/anonymization-iii-risk-re-identification.

[2] AEPD and EDPS. 10 misunderstandings related to anonymisation. Guidance document, European Data Protection Supervisor and Spanish Data Protection Agency, 04 2021.

[3] Article 29 Data Protection Working Party. Opinion 05/2014 on anonymisation techniques. Technical Report WP216, European Commission, 4 2014. 0829/14/EN.

[4] Andrew Burt, Sophie Stalla-Bourdillon, and Alfred Rossi. A guide to the eu's unclear anonymization standards, 2021. Accessed August 10, 2025. https://iapp.org/news/a/a-guide-to-the-eus-unclear-anonymization-standards.

[5] CNIL. Decision n°san-2023-00 Doctissimo, 2023.

[6] Aloni Cohen and Kobbi Nissim. Towards formalizing the gdpr's notion of singling out. *Proceedings of the National Academy of Sciences 117(15), 8344–8352*, 2020.

[7] Court of Justice of the European Union. Case C-582/14 Breyer, 10 2016.

[8] Court of Justice of the European Union. Case C-319/22 Scania, 11 2023.

[9] Court of Justice of the European Union. Case C-479/22 Olaf, 09 2024.

[10] Court of Justice of the European Union. Case C-604/22 IAB, 3 2024.

[11] Court of Justice of the European Union. Case C-413/23p SRB, 09 2025.

[12] Lorenzo Dalla Corte. coping personal data: Towards a nuanced interpretaton of the material scope of eu data protecton law. *European Journal of Law and Technology, Vol 10, Issue 1*, 2019.

[13] Data Protection Commission Ireland. Guidance on anonymisation and pseudonymisation, 2019.

[14] DROPS Project. DROPS Research Project, 2025. Accessed August 10, 2025. https://itsec.cs.uni-bonn.de/drops/en/.

[15] Khaled El Emam and Cecilia Alvarez. A critical appraisal of the article 29 working party opinion 05/2014 on data anonymization techniques. *International Data Privacy Law 5(1), 73–87*, 2015.

[16] Samson Yoseph Esayas. The role of anonymisation and pseudonymisation under the eu data privacy rules: beyond the 'all or nothing' approach. *in European Journal of Law and Technology, Vol 6, No 2*, 2015.

[17] European Data Protection Board. Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models, 2024.

[18] European Data Protection Board. Guidelines on pseudonymisation, 2025.

[19] European General Court. Case T-557/20 SRB, 2023.

[20] European Medicines Agency. European medicines agency policy on publication of clinical data for medicinal products for human use. Policy EMA/144064/2019, European Medicines Agency, 3 2019.

[21] Michèle Finck and Frank Pallas. They who must not be identified—distinguishing personal from non-personal data under the gdpr. *International Data Privacy Law*, 10(1):11–36, 03 2020.

[22] Andrea Gadotti, Luc Rocher, Florimond Houssiau, Ana-Maria Creţu, and Yves-Alexandre De Montjoye. Anonymization: The imperfect science of using data while preserving privacy. *Science Advances 10(29), eadn7053*, 2024.

[23] Garante. Decision PDP [9913795] Registro dei provvedimentin°226 del 1° giugno 2023 1 The THIN database, 2023.

[24] Mostafa Langarizadeh, Azam Orooji, and Abbas Sheikhtaheri. Effectiveness of anonymization methods in preserving patients' privacy: a systematic literature review. *Health Informatics Meets eHealth*, pages 80–87, 2018.

[25] Szivia LestyĂAn, William Letrone, Ludovica Robustelli, and Gergely BiczĂłk. Anonymity-washing. *arXiv preprint arXiv:2505.18627*, 2025.

[26] Alexandre Lodie and Cédric Lauradoux. Is it personal data? solving the gordian knot of anonymisation. In *Privacy Symposium, Venice*, 2024.

[27] Csányi Gergely Márk, Daniel Nagy, Renátó Vági, Pal Vadász, and Tamás Orosz. Challenges and open problems of legal document anonymization. *Symmetry, 13(8):1490*, 2021.

[28] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *IEEE SP*, 2008.

[29] Paul Ohm. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review 57, 1701*, 2009.

[30] OWASP, Cheat Sheets Series Team. Password storage cheat sheet, 2025. Accessed August 10, 2025. https://cheatsheetseries.owasp.org/cheatsheets/Password_Storage_Cheat_

Sheet.html.

[31] Nadezhda Purtova. From knowing by name to targeting: the meaning of identification under the gdpr. *International Data Privacy Law, 12*, 2022.

[32] Ira S. Rubinstein and Woodrow Hartzog. Anonymization and risk. *Wash. L. Rev. 91, 703*.

[33] Valentin Rupp and Max von Grafenstein. Clarifying "personal data" and the role of anonymisation in data protection law: Including and excluding data from the scope of the gdpr (more clearly) through refining the concept of data protection. *Computer Law Security Review, Volume 52*, 2024.

[34] Rolf Schwartmann, Andreas Jaspers, Niels Lepperhoff, Steffen Weiß, and Michael Meier. Practice guide to anonymising personal data: Requirements, application classes and procedure model. practice guide, Foundation for Data Protection. Original title: Praxisleitfaden zum Anonymisieren personenbezogener Daten.

[35] Gerald Spindler and Philipp Schmechel. Personal data and encryption in the european general data protection regulation. *IPITEC*, 2016.

[36] Sophie Stalla-Bourdillon. Identifiability, as a Data Risk: Is a Uniform Approach to Anonymisation About to Emerge in the EU? pages 1–19.

[37] Sophie Stalla-Bourdillon and Alison Knight. Anonymous Data v. Personal Data - False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data. 34:284.

[38] Latanya Sweeney. k-anonymity: A model for protecting privacy. *IJUFKS 10(5)*, page 557–570, 2002.

[39] Kalliopi Terzidou. Automated anonymization of court decisions: Facilitating the publication of court decisions through algorithmic systems. In *n Nineteenth International Conference for Artificial Intelligence and Law (ICAIL 2023)*, 2023.

[40] Emily Weitzenboeck, Pierre Lison, Malgorzata Agnieszka Cyndecka, and Malcolm Langford. The GDPR and Unstructured Data: Is Anonymization Possible?