

Anonymity-washing

Szilvia Lestyán^{1,2}, William Letrone³, Ludovica Robustelli³
Gergely Biczók⁴[0000–0002–3891–3855]

¹ Inria

² Institut national d’études démographiques, France
`szilvia.lestyan@inria.fr`

³ DCS, Nantes University, CNRS

`{william.letrone,ludovica.robustelli}@univ-nantes.fr`

⁴ CrySyS Lab, Budapest Univ. of Technology and Economics
`biczok@crysys.hu`

Abstract. Anonymization is a foundational principle of data privacy regulation, yet its practical application remains riddled with ambiguity and inconsistency. This paper introduces the concept of anonymity-washing—the misrepresentation of the anonymity level of “sanitized” personal data—as a critical privacy concern. While both legal and technical critiques of anonymization exist, they tend to address isolated aspects of the problem. In contrast, this paper offers a comprehensive overview of the conditions that enable anonymity-washing. It synthesizes fragmented legal interpretations, technical misunderstandings, and outdated regulatory guidance and complements them with a systematic review of national and international resources, including legal cases, data protection authority guidelines, and technical documentation. Our findings reveal a lack of coherent support for practitioners, contributing to the persistent misuse of pseudonymization and obsolete anonymization techniques. We conclude by recommending targeted education, clearer technical guidance, and closer cooperation between regulators, researchers, and industry to bridge the gap between legal norms and technical reality.

1 Introduction

Anonymization is widely regarded as a crucial tool for protecting privacy in an era of big data processing. Theoretically, it serves as a means to mitigate risks associated with the misuse of personal data by ensuring that individuals can no longer be identified. In practice, however, anonymization remains an imprecise science, often misunderstood and misapplied. Many datasets that are presented as anonymized continue to pose significant re-identification risks due to improper techniques or evolving technological capabilities. This gap between the intended function of anonymization and its real-world implementation has led to growing concerns about *anonymity-washing* — a phenomenon in which organizations claim to have achieved strong privacy protections through anonymization while

failing to provide meaningful safeguards. Note that anonymity-washing is a specialized form of *privacy-washing*⁵ [19].

The General Data Protection Regulation (GDPR) establishes anonymization as a mechanism through which personal data can be rendered outside the scope of data protection laws. Recital 26 of the GDPR defines anonymization as the process by which data is “*rendered anonymous in such a manner that the data subject is not or no longer identifiable*.” However, the absence of clear, practical guidance on how to achieve this standard has resulted in inconsistent implementations and legal uncertainties. Many organizations either overestimate the effectiveness of their anonymization processes or struggle to comply due to conflicting regulatory interpretations. Additionally, courts have recognized that anonymization is never absolute — what is considered anonymous today may become identifiable tomorrow as technology advances. Despite the importance of anonymization, the regulatory and educational landscape remains fragmented and inadequate. On one end of the spectrum, legal guidelines provide high-level definitions and compliance requirements but lack technical specificity. On the other end, academic research offers rigorous, mathematically grounded approaches to anonymization that are often inaccessible to practitioners who do not have advanced expertise in statistics or computer science. This disconnect has left engineers, data scientists, and policymakers without the necessary tools to implement anonymization effectively. The result is widespread reliance on outdated or insufficient methods—such as k-anonymity and l-diversity—that have been repeatedly shown to fail against modern re-identification attacks [46].

Furthermore, anonymity-washing is exacerbated by inconsistent regulatory interpretations across jurisdictions. The European Union has exercised significant global influence on data privacy regulation, with many countries modelling their laws after the GDPR. However, even within the EU, national data protection authorities and courts have issued conflicting opinions on what constitutes effective anonymization, leading to uncertainty among organizations attempting to comply. Beyond Europe, frameworks such as the United States’ de-identification standards under the Health Insurance Portability and Accountability Act (HIPAA) and the California Consumer Privacy Act (CCPA), Japan’s Act on the Protection of Personal Information (APPI), and emerging guidelines such as the Brazilian General Data Protection Law (LGPD) further demonstrate that approaches to anonymization lack uniformity at the international level, making cross-border data governance exceedingly complex.

Another critical factor enabling anonymity-washing is the lack of accessible educational resources for practitioners. Engineers and software developers responsible for implementing anonymization frequently lack adequate training and rely on either high-level legal guidelines or complex, research-oriented papers that do not offer practical guidance. Several regulatory bodies and experts have called for clearer standards, including the European Data Protection Board (EDPB), national data protection authorities (such as the National Commission

⁵ a particularly timely research theme, see <https://www.dagstuhl.de/en/seminars/seminar-calendar/seminar-details/25112>

for Information Technology and Civil Liberties in France (CNIL) and the Federal Commissioner for Data Protection and Freedom of Information (BfDI) in Germany), and research institutions. Yet, despite these calls for action, practitioners continue to report difficulties in accessing concrete, actionable information on how to apply anonymization techniques effectively.

In light of these challenges, this paper argues that ambiguities in regulatory guidance, outdated technical approaches, and gaps in practitioner education may lead to anonymity-washing. While prior works have addressed specific aspects of the problem—such as legal critiques of anonymization under data protection law [83,90,75,16], technical limitations of anonymization techniques [1,24,65,46,40,26], or even highlighting key misunderstandings [37]—these contributions offer only a partial view of the broader landscape. In contrast, our work provides a comprehensive analysis of the multiple, interrelated issues underlying anonymity-washing. We expand on the existing literature by integrating a wide range of sources, including legal cases, regulatory interpretations, and technical guidelines, while offering a systematic critique of technical documentation. Furthermore, we provide an international perspective that, to our knowledge, has not been previously compiled in a single work.

First, in Section 2 we introduce the concept of anonymity-washing and situate it within the broader landscape of privacy discourse. In Section 3, we examine the legal foundations and the regulatory ambiguity surrounding anonymization. Next, Section 4 presents an overview and critique of technical guidelines and educational resources, highlighting the gaps practitioners face. Section 5.1 explores the practical implications of anonymity-washing, including legal cases and implementation failures. Finally, Section 6 offers recommendations and concluding reflections on how to address the risks of anonymity-washing through clearer guidance and improved institutional coordination.

2 Contextual Elements

2.1 Anonymization terminology

The anonymization landscape is complex, with multiple laws advocating for different requirements. But many points of contention stem from the terminology surrounding the topic of anonymization. To begin, it is interesting to look at the terminology developed by the International Organization for Standardization (ISO), as it constitutes the main standard-setting body with international influence. Several ISO standards touch on the topic of data anonymization. These global standards have recognized the importance of anonymization in various contexts. ISO/IEC 29100:2024(en) establishing a common privacy terminology, defines anonymization as a

“[A] process by which personally identifiable information (...) is irreversibly altered in such a way that a [data subject] (...) can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party.”

The same document defines pseudonymization as a

“[A] process applied to personally identifiable information (PII) (3.7) which replaces identifying information with an alias.”

The other term, “de-identification”, is usually considered more neutral and broader than anonymization, although sometimes conflated with the latter [18,47]. Indeed, according to ISO, “de-identification” refers to

“[A] process of removing the association between a set of identifying attributes (3.14) and the data subject (3.4).”⁶

It results that anonymization implies the highest degree of privacy, while the more specific process of pseudonymization is a step below anonymization in terms of re-identifiability. In contrast “de-identification” is the general term describing the process through which data is made confidential⁷. While some jurisdictions mostly follow the ISO terminology, others, unfortunately, do not [2,94]. An example is the fact that the term “de-identification” is not even used within the EU’s GDPR, while several important US instruments, such as the HIPAA [93] and the CCPA [17] use it in place of anonymization. In the same vein, Nigeria and Malawi’s Data Protection Acts do not use the term “anonymization”, despite referring to both “de-identification” and “pseudonymization” in their statutes [2]. In contrast, Japan’s *Act on the Protection of Personal Information*, much like the EU, does not refer to de-identification. Finally, a cursory look at the relevant literature in social science reveals that authors themselves appear to have subscribed to different terminologies [46,18].

Beyond word choice, there seems to be no equivalence between the terms when they are used to refer to data records that have undergone the appropriate treatment to exempt data controllers and processors from their obligations under data protection laws. That is to say, the tolerance level towards identifiability tends to vary across jurisdictions [2]. Discrepancies sometimes exist within a single legal system, as in the US, where re-identifiability tolerance may vary depending on the nature of the data contained in a record, and the projected use of the record [47]. In the EU, the situation is no less confusing, as “anonymization” suffers from conflicting interpretations [40] (see details in Section 3).

Take-away

These variations and inconsistencies make it difficult for practitioners to understand and determine the required level of protection, hindering the understanding and adequate application of anonymization techniques.

⁶ ISO/IEC 20889:2018(en) Privacy enhancing data de-identification terminology and classification of techniques. See also, the more recent ISO/IEC 5207:2024(en) Information technology — Data usage — Terminology and use cases.

⁷ Note that the technical terminology could be used slightly differently; here we are discussing only the legal definitions.

2.2 “Anonymity-washing” as the misrepresentation of actual confidentiality levels

Due to interpretative instability, the terms that compose the anonymization terminology should not necessarily be taken at face value. Not only are practitioners affected by the confusion in the terminology, but individuals are affected as well, as they may put more trust in information processes than they should. In order to better understand this effect, we must look at interpretations of *privacy-washing*. In the course of an analysis on questionable data practices of tech industry giants, Girucci gives the following definition [19]:

“The purposeful conflation of security with privacy, the disregarding of more granular definitions of privacy (social vs. institutional privacy as well as data types including explicit, implicit, aggregated, and inferred), and a general reliance on offline privacy expectations that are no longer applicable to online spaces.”.

Despite its provocative tone, the term privacy-washing is more than a mere rhetorical device. Indeed, privacy-washing can accurately describe situations where data privacy guarantees deviate from the standards to which the concerned entities purportedly committed. Evidently, the concept of privacy-washing is broad: it can cover a variety of subjects like cybersecurity and third-party data sharing. This paper is focused on privacy-washing in the anonymization context because deceptive privacy representations in this context are highly likely. In fact, while the anonymization vocabulary taken at face value is unambiguous, it does little to convey the actual fragility [75] of current anonymization methods:

“The way companies and the media talk about de-identified data matters, and data holders regularly play fast and loose with the concept of anonymity. The terms “anonymous” and “anonymization” simply over-promise. They create expectations of near-perfection and lull people into a false sense of security” [83].

Data controllers could be tempted to exploit the complexity within current data privacy terminology to mislead data subjects regarding the safety and confidentiality of their data, resulting in anonymity-washing. In essence, anonymity-washing refers to situations involving the misrepresentation of anonymity levels of a data record. Recently, the Federal Trade Commission (FTC), as the main agency dealing with consumer protection in the US, dealt with anonymity-washing cases. In a recent communication, the FTC warned that unwarranted claims of anonymity could constitute deceptive consumer practices, reiterating that pseudonymous identifiers in the form of hashing do not constitute anonymization, as some businesses have claimed:

“Companies should not act or claim as if hashing personal information renders it anonymized. FTC staff will remain vigilant to ensure companies are following the law and take action when the privacy claims they make are deceptive”[25].

Remark, how this highlights the manipulative aspect of anonymity-washing.⁸ In the EU, potential anonymity-washing cases have been scrutinized by data protection authorities, and some practices have been challenged in Court. For instance, the Italian Data Protection Authority (Garante) recently sanctioned the Italian National Institute of Statistics for its failure to deploy the necessary measures to avoid re-identification of the data it used for statistical analysis. In its order, the Italian authority explained [48];

“Simply having organizational measures or ethical codes is not enough to satisfy data protection principles.”

In this case, data controllers claimed to have upheld data protection principles while the data subjects remained, in fact, easily re-identifiable from their data records.

The question of intentionality behind deceitful anonymity statements deserves a brief focus, as the term “washing” implies an intentional action. Except that intention in this context can be difficult to prove. Sometimes, anonymity-washing cases are so blatant that the willingness to deceive leaves no doubt. Other times, anonymity-washing is harder to prove and therefore appears incidental, giving the impression that data controllers and/or processors are acting in good faith while deploying weaker solutions. There is, of course, a risk of mischaracterization. Still, it may never be possible to prove with a high degree of confidence that a data controller and/or processor acted in good faith, since defendants are likely to claim to be acting in good faith when notified, and in the course of legal proceedings.

Take-away

Anonymity-washing is a subset of privacy-washing, which refers to the misrepresentation of the anonymity level of data. The phenomenon is exacerbated by several factors, including unclear terminology.

3 Overview of regulatory guidance on data anonymization

On 25 July 2024, the European Commission published its second report on the implementation of the GDPR [44]. One of the key issues highlighted in the report is the persistence of differing interpretations among national data protection authorities, which undermines the uniform application of the GDPR. This discrepancy gives rise to legal uncertainty, thus businesses are confronted with divergent administrative requirements across different Member States. In this regard, the Commission seeks to reiterate its request, previously made in

⁸ It bears noting that the FTC’s view on hashing is consistent with the practices in the EU, where hashing constitutes a method of pseudonymization that does not suffice on its own, in making data records fall outside the GDPR’s scope, due to the likelihood of privacy harms that may result from sharing the records.

2020 [35], to support practitioners by providing clearer guidance and materials to facilitate GDPR compliance. This issue is particularly pertinent in the context of anonymity washing, as the Commission has reported in [36]:

“Some stakeholders also consider that certain data protection authorities and the Board adopt interpretations that deviate from the risk-based approach of the GDPR, [and] (...) mention as areas of concern: (i) the interpretation of anonymization; (...)” [As a result, the report] *“underline[s] the need for additional guidelines, in particular on anonymization and pseudonymization (...)”*.

3.1 EU regulations

The abrogation of Directive 95/46/EC [33] (Data Protection Directive or DPD) and the adoption of the GDPR did not affect anonymization. This is confirmed by the endorsement of the Working Party 29’s (WP29) Opinion 5/2014 on anonymization by the European Data Protection Board (EDPB), which is still in the process of preparing an updated version [88]. In its Opinion 5/2014 on anonymization [78], WP29 recalls the ISO definition of anonymization⁹ and that the simple removal of identifiers from personal data does not make the anonymization process irreversible. Account should be taken of all “reasonable means” (including computational power and technological evolution) to re-identify anonymous data. These points are addressed by Recital 26 of the GDPR, stating that :

*“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. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.”*¹⁰

Before deciding on an anonymization method, an anonymization test must be performed to evaluate the risks (singling-out, linkability, and inference). WP29 provides an assessment of the guarantees and shortcomings of each technique from the two main families of anonymization (generalization and randomization) based on these risks. However, several research papers have shown that the analyses provided by WP29 have weak points, and they do not consider these techniques valid [24,84,69,29,87,70]¹¹.

⁹ ISO 29100:2024

¹⁰ The previous directive 95/46/EC, which the WP29 Opinion 5/2014 referred to, was not as detailed. It contained a general reference to reasonable means that are likely to be used to re-identify the data (without including the notion of ‘singling out’) and relied on codes of conduct to encourage anonymization practices.

¹¹ For example, the k-anonymity technique does not prevent the risk of singling out, contrary to the conclusions of WP29.

More recent guidelines have been adopted by the EDPB on issues related to data anonymization, but they do not contain additional advice. On 17 December 2024, the EDPB published guidelines on the anonymization of AI models [38]. It states that whenever models are trained on personal data, they cannot be considered anonymous. The reason is that many studies on these models have demonstrated their capacity to “regurgitate” part of their training datasets [6]. The EDPB stated that a model is considered anonymous only when, based on appropriate documentation, personal data cannot be inferred either directly (by statistical inference, including the probabilistic functioning of the model) or indirectly (within a user’s prompt). If the risk of “regurgitation” of personal data persists, a deeper analysis is needed¹².

Sénéchal criticizes the lack of a threshold of the risk of “regurgitating” personal data and the lack of a distinction between the different AI models in these guidelines in [92] (for example, the general-purpose AI models [76] and the ones posing systemic risks. This is problematic since anonymization is difficult to implement, especially with unstructured data [95], which are essentially used to train general-purpose AI models [96]. Additionally, the question of whether the data can be separated from the model remains unanswered. It is also not known whether the anonymization of the model implies that of the data it contains.

The EDPB also adopted on 16 January 2025, guidelines on pseudonymization [39], in which it recalls the GDPR definition set out in article 4(5). The Board stressed that, although pseudonymization secures data, whenever the re-attribution of data to a natural person (by linking pseudonyms to additional data) remains possible, the GDPR applies. It recalls that even if the original data are deleted, pseudonymized data become anonymous only if all requirements are met. It is interesting to note that the *guidelines do not provide further information on anonymization requirements*. This is regrettable for two reasons: first, updated guidelines on anonymization have yet to be issued; and second, as the Spanish Data Protection Authority has pointed out, confusion between anonymization and pseudonymization remains a common misunderstanding among data controllers [37].

Take-away

Guidelines on anonymization need to be updated (as they have not been since 2014). The information provided by the EDPB on pseudonymization and anonymization of AI models does not resolve the contradictions of its previous guidelines and the practical difficulties controllers are confronted with when implementing anonymization protocols in real life.

¹² The following aspects are required for verification: the source of the data, their preparation and minimization, the training method, the analysis of the model, the resistance of the model to cyber attacks and the documentation provided.

3.2 Anonymization regimes beyond the EU

The uncertainties resulting from the changing interpretation within the EU undermine the so-called “Brussels effect”, when non-EU states take inspiration from the EU’s laws for building their own legal regime. Data flows often involve entities located in different jurisdictions, including non-EU countries [68,67]. Moreover, data protection laws usually have some extraterritorial effects, which means that multiple regimes are sometimes applicable simultaneously. Hence, it is vital to ensure that legal regimes on anonymization do not contradict each other. Yet, a survey conducted by the OECD in 2019 found that

“*uncertainty regarding legal privacy regimes*” and “*incompatibility of legal regimes*” topped the list of the main challenges to cross-border data flows [74].

Anonymization guidelines are present in data protection regimes across the globe. There are differences, however, in the approaches and the overall granularity levels exhibited by the relevant frameworks. Notably, some data protection regimes, such as in Japan [58] and the US [93], come with relatively detailed guidance on how to achieve the expected levels of anonymization and how to handle the data [85,5]. With regard to data transfer between the EU and the USA the previous Privacy Shield, which was adopted on the basis of the European Commission’s decision that the USA’s level of personal data protection was equivalent to that of the EU, [42], was replaced by a revised Privacy Framework, after the EU Court of Justice overturned it [60]. In contrast, other regulatory frameworks, such as in Brazil [14], are not particularly prescriptive and require additional input. At the same time, several jurisdictions have initiated efforts to modernize their approaches to data protection, including anonymization. For example, Brazil’s national data protection authority, the ANDP, is set to clarify what measures could be implemented to ensure anonymity in accordance with its 2018 *Lei Geral de Proteção de Dados* (LGPD) in the upcoming years [4]. A call for public participation in that effect has been published in early 2024. In 2023, the Data Security Council of India (DSCI) published a roadmap considering possible orientations for a national data anonymization regime [53]. At the intra-state level, in Québec, the *Regulation respecting the anonymization of personal information* was published in 2024 [82]. The text is very prescriptive and seeks to clarify the distinction between anonymity and pseudonymity, aligning with the EU’s view. In the EU, the EDPB is expected to publish new guidelines on anonymization later this year. It is expected that the new guidelines will fix the inconsistencies introduced by the WP29 Opinion 5/2014 on anonymization, thereby clarifying the dominant approach at the EU level [40].

Anonymization is still a maturing field. Valuable guidelines on anonymization are often released after the publication of the main body of law. Hence, recent data protection laws such as China’s *Personal Information Protection Law*, or India’s *Digital Personal Data Protection Act*, will need to be complemented with guidelines on anonymization [89,68]. Furthermore, while precise anonymization parameters are still not consistent across jurisdictions, the basic premises of

anonymization law remain the same; anonymization levels may vary, and so are the obligations placed upon data handlers [68].

Whatever the approach, it seems that regulators are left with two choices: either leaving enough leeway for data handlers to determine for themselves which methods and policies would meet their expectations, or prescribing exactly which technical and organizational measures would meet their expectations. Both approaches have their merits and shortcomings. On the one hand, there is an inherent limitation on the degree of granularity that can be achieved in the law. Excessively precise regulations and guidelines may pose problems at the implementation stage and may prove to be overly restrictive, as has already been seen with the WP 29 Opinion 5/2014. The limited technical knowledge of regulators may constrain the formulation of highly detailed guidelines anyway. On the other hand, too much leeway could seriously undermine the purpose of data protection laws by increasing the likelihood that poorly anonymized data records will fall outside their scope.

Take-away

Anonymization laws seem to have been evolving independently, with differing requirements and definitions. Whether all the ambiguities will be fixed and whether every actor will converge around the same interpretation remains to be seen.

4 Overview and Critique of Guidelines

4.1 Contradictory guidelines and uncertain standards within the EU

Some authors consider that the EU Data Protection Law lacks a clear definition of anonymization [16]. Unlike pseudonymization, the GDPR fails to define anonymization in its Article 4 titled “definitions”. However, another Opinion of WP29 on the concept of personal data issued in 2007 [79] clarified the difference between anonymization and pseudonymization. Recalling ISO’s previous definitions, they explained that anonymization protects privacy, while pseudonymization represents a technical, reversible process. Nevertheless, data can still be considered anonymous, even when re-identification remains possible, but complementary measures to prevent re-identification are implemented. This flexible approach was not supported in the Opinion 5/2014 on anonymization [78], which applies together with the previous Opinion on the concept of personal data. In Opinion 5/2014, WP29 required the aggregation of data (into group statistics) and the destruction of raw data (identifiers) to ensure correct anonymization. Nevertheless, the objective still remained to prove that the likelihood of re-identification was negligible. On this point, the Commission’s guidelines [34] on the free flow of non-personal data suggest that it is often difficult to assess the effectiveness of an anonymization procedure. Indeed, besides many academic

papers [3,46,29,70,1,28], even a study commissioned by the European Parliament’s ITRE Committee has shown that it is possible to re-identify supposedly anonymized data [45].

Moreover, national DPAs disagree on how to implement anonymization. The French DPA, CNIL, adopts the WP29’s approach to anonymization; other national DPAs are more flexible. For example, the UK’s ICO (Information Commissioner’s Office) states that [50]:

“The DPA does not require anonymization to be completely risk free—you must be able to mitigate the risk of identification until it is remote. If the risk of identification is reasonably likely, the information should be regarded as personal data— (...). Clearly, 100% anonymization is the most desirable position, and in some cases, this is possible, but it is not the test the DPA requires.”,

and Ireland’s DPC (Data Protection Commission) writes[32]:

“Organisations don’t have to be able to prove that it is impossible for any data subject to be identified in order for an anonymization technique to be considered successful. Rather, if it can be shown that it is unlikely that a data subject will be identified given the circumstances of the individual case and the state of technology, the data can be considered anonymous”.

The CNIL’s guidelines on anonymization are of particular interest, as they establish a strict standard to determine whether data can be considered anonymous. They assert that data is anonymous only when it is *impossible* to re-identify the data subject. However, they recognise that when the risks of singling out, linkability, and inference are not met, data can be deemed anonymous if a subsequent analysis indicates a negligible risk of re-identification [21]. Unfortunately, the definition of what constitutes a *negligible* risk remains ambiguous.

For these reasons, many controllers have given up on anonymization and prefer pseudonymization [16] even if pseudonymized data remain subject to the GDPR. The problem in this case is that *“pseudonymized data are considered personal data, regardless of whether they are, or ever will be, in the hands of a person who holds the key needed for re-identification”* [81]. The application of personal data regulation can be challenging, especially in the context of rapid data access, as evidenced by the case of health data (see also Section 5.1).

Take-away

The lack of a clear, harmonized definition of anonymization across EU legal texts and among national authorities creates confusion for practitioners. This leads many to favour pseudonymization, even though it offers weaker privacy protection, with data remaining fully subject to GDPR. This regulatory ambiguity undermines the consistent implementation of and weakens trust in anonymization as a reliable privacy safeguard.

4.2 Technical Documents

While non-technical guidance often lacks the precision needed for implementation, technical documentation is not always more helpful. In several cases, companies have claimed they could not find clear guidance on how to anonymize data—a claim sometimes countered by DPAs pointing to existing documents [23]. However, our review shows that most guidelines are often hard to find (not available or ignored by DPAs) or not practically useful (entry-level). We reviewed the websites of the five most active EU DPAs (France, Austria, Ireland, Germany, and Italy). Most do not provide detailed technical materials:

- CNIL (France) offers introductory guides [20,21], repeating WP29 content, and a clear (though potentially misleading) explanation of pseudonymization [22].
- Garante (Italy) provides an overview [49] to implement the GDPR, and mainly reiterates previous legal guidelines.
- BfDI (Germany) has policy papers and speeches [7,8,11,12], but limited technical depth. [15] focuses on the importance of anonymization, [9] and [10] discuss risks of other issues related to personal data.
- DSB (Austria) offers legal advice only.
- DPC (Ireland) stands out with well-structured and clear guidance [31,32] on legal questions, however, it does not offer practical advice.

Outside the EU, the UK’s ICO provides excellent guidance, including on state-of-the-art methods like differential privacy and other PETs [51,52]. However, some examples are oversimplified or technically wrong¹³

The UK Anonymization Network (UKAN) also offers practical tools, such as a decision-making framework that uniquely addresses attacker modelling [41]. However, most chapters remain general (entry-level), in contrast to the referenced DIS method that requires Master’s degree-level statistical knowledge.

The anonymization guide of Singapore’s PDPC [80] is accessible and educative, guides the reader from data discovery to risk measures, giving informative examples; however, it is also entry-level and adds little beyond other existing material.

Some statistical agencies provide additional resources. The National Institute of Statistics and Economic Studies in France (INSEE) offers slides and working papers [54,56], but most lack practical detail. A notable exception is Bergeat’s work [57] comparing and explaining experiments done by two anonymization software tools: μ -Argus and SDCMicro. It also gives plenty of citations, however, only to sources on statistics (no computer science references). It is aimed at statisticians, and it could serve as a continuation to other introductory materials, but the reader who already has at least a Bachelor’s degree in mathematics or statistics. Other guides, such as [55], focus on confidentiality rules within the French statistical service rather than techniques. Statistical documents only mention statistical tools and use a different language from that of computer scientists.

¹³ See the case study on differentially private mixed noise addition

This could, unfortunately, result in ignoring some state-of-the-art methods, such as differential privacy. A good example is the working paper [30] that details the anonymization process applied to a large French administrative database where the authors experimented with different methods including k-anonymity, all-m anonymity, and l-diversity. They mention that they have tried to apply DP, however, they abandoned the experiment due to a lack of expertise.

Academic papers are another option, but they often assume advanced statistical or mathematical knowledge (Master’s or PhD level), making them inaccessible to many practitioners. Moreover, choosing appropriate methods from the literature is difficult without deep expertise, which may explain the frequent use of outdated or misapplied techniques in practice [56,57,66,30]. Some expert-written materials aimed at non-technical readers exist [97,72,71]; however, they are rarely cited in public or institutional guidance.

Books Books on anonymization tend to target either high-level management (e.g. [27,86,73]) or technical researchers. Some, like [13], cover a broad range of privacy topics but lack methods for evaluating anonymization quality. Jar-mul’s work [59] offers a more hands-on perspective, including differential privacy and privacy engineering workflows, making it useful for practitioners. Stallings’ book [91] is a strong general-purpose resource, well-suited for short training programs.

Take-away

Most technical anonymization resources are either too simplistic or too involved, offering little practical use for professionals. Practical regulatory guidance is rare and often legalistic. This leaves practitioners with a fragmented landscape, outdated methods, and an incentive to abandon anonymization altogether. Bridging these gaps requires targeted, accessible, and technically sound educational materials.

5 Inadequate Practices

5.1 Confusions arising from the definition of personal data

EU case law lacks clarity with regard to anonymization practices. The most relevant cases focus on clarifying the concept of personal data. However, the interpretations provided help to assess what an anonymized dataset is *not*. Moreover, the Court’s application of Recital 26 to real cases provides valuable insight into the question of whether data remains anonymous despite a residual risk of re-identification. In this regard, the General Court’s SRB vs EDPS decision is a good example [43]. SRB (Single Resolution Board) carried out an insolvency procedure against Banco Popular. Within this procedure, some data were processed to assess the eligibility of the participants for compensation. Each participant was identifiable by means of an alphanumeric code generated randomly. The staff

processing this data only had access to these codes and not to the key identifying them. The EDPS (European Data Protection Supervisor) considered these data pseudonymized [77], but its decision was challenged before the General Court. Using a risk-based approach, the General Court decided that the data were anonymous. Indeed, according to the Breyer Court of Justice case law [61], the additional information (the key) needed to re-identify the data subjects remained inaccessible to the processing staff. The fact that the staff could not legally access the complementary data that would allow re-identification proved enough to consider that no *reasonable means* existed to re-identify the data, which thus remained non-personal [43].

5.2 EUCJ’s case-law on personal data

This example shows that the EUCJ’s case law on personal data builds upon its precedents rather than undergoing a radical evolution. On this point, Breyer’s decision [61] was about the dynamic nature of IP addresses. These IP addresses are subject to change with each connection. The plaintiff initiated legal proceedings against the Republic of Germany for its practices concerning the storage and registration of this data. The Court of Justice had to determine whether dynamic IP addresses should be considered personal data for the service provider. The Court decided that the retention of all information by a single individual was not a prerequisite for data being considered as personal. This meant that a third party could retain such re-identifying information, and that this circumstance did not affect the qualification of the data. However, the Court acknowledged that an assessment was necessary to determine the reasonableness of combining this information, taking into account the effort, time, and cost associated with the operation, as well as the accessibility of this additional information (enabling user identification) to the service provider. Given the legal restrictions on such access in Germany, the Court determined that in the absence of legal means to obtain this information, the data in question were not deemed personal. The doctrine posits that two fundamental elements have been applied since the Breyer decision to ascertain the personal nature of data. These are: (1) the distinguishability of the data, defined as the capacity of the data points to identify an individual, and (2) the availability of additional data to “situationally relevant entities” who are capable of associating these data with a physical person [89].

The Scania decision perfectly [64] illustrates this methodology. In this case, the Court was asked to determine the legal status of a vehicle identification number (VIN), a unique alphanumeric code assigned by manufacturers to identify the proprietor of a vehicle. In its decision, the Court stated that the VIN can be personal data for independent operators and vehicle manufacturers if the former have the additional data that enable re-identification, and for the latter if they make the VIN available. The availability of data is considered in conjunction with the capacity of isolating proprietors of the vehicles or all other people who have a title on them. Some authors have suggested an evolution in the interpretation of personal data, attributing this change to the Court’s categorization

of independent operators and manufacturers as “situationally relevant entities”, capable of associating VIN with additional identifying information.

Take-away

The consistency of the Court’s jurisprudence on the concept of personal data is paramount to contrast the phenomenon of anonymity-washing. This is particularly crucial given the occurrence of poor anonymization practices, which the EUCJ is entitled to sanction.

5.3 Confusion at the institutional level

A relevant case law that sanctioned the European Commission [62] demonstrates the importance of taking into account publicly available data to assess the risk of re-identification. The applicant received European funding as a researcher. The funds had been misappropriated, and the costs were ordered to be reimbursed. The Commission published a press release summarizing the decision, without mentioning the applicant’s direct identifiers in order to protect their privacy. However, the researcher brought an action for the annulment of the press release, since it contained identifiable data. The General Court dismissed it, and the applicant appealed its decision to the EUCJ. The Court considered that *“information relating to the gender of a person who is the subject of a press release, that person’s nationality, his or her father’s occupation, the amount of the grant for a scientific project and the geographical location of the entity hosting that scientific project, taken together, contain information that may allow the person who is the subject of that press release to be identified, in particular by those working in the same scientific field and familiar with that person’s professional background”* and goes on that this circumstance, *“does not allow the risk of identification of the data subject to be regarded as insignificant.”* The judgment avoids discussing anonymization practices in detail but highlights that this issue is subject to misinterpretations, including among public institutions like the Commission. In fact, deleting direct identifiers seems sufficient to achieve anonymization, even by the EU institutions.

5.4 Confusion within firms

Confusion about anonymization practices is also widespread among companies. A good example is the IAB Europe case-law [63]. The company established a set of guidelines aimed at ensuring compliance with the GDPR concerning the collection of browsing data via a TC String (a series of characters coding the user’s preferences). This string could later be used by companies for commercial purposes. The Court ruled that the TC String was a form of personal data, given its capacity to allow individuals to be identified by associating it with additional information (such as an IP address). Despite third parties retaining this additional information, IAB Europe was able to obtain it.

In this regard, the relationship between the EUCJ case-law, employing a risk-based approach to assess the reasonable means likely to be used to re-identify the data, and the WP29 Opinion 5/2014 on anonymization appears complex [90]. On the one hand, discordance persists in the discourse of WP 29 between the zero-risk approach (re-identification must be negligible if not impossible) and the necessity for reasonableness, given that all anonymization techniques are considered imperfect [40]. On the other hand, the need to destroy raw data and to aggregate them in order to achieve anonymization, as required by WP 29, is not met by the case law of the EUCJ, which considers data to be anonymous even if the original data are not deleted, the only relevant aspect being the impossibility (legal rather than technical) to access the additional data that enable re-identification. These contradictions contribute to privacy washing practices by making it difficult to distinguish the company’s bad faith from its lack of knowledge about anonymization methods, especially when data are processed for commercial purposes. In the case of IAB Europe, for example, the company assumed that due to the unavailability of additional information, TC String did not constitute personal data. This mistake is frequently observed among firms. This assertion is supported by numerous decisions adopted by the national DPAs.

5.5 Confusions arising from the difference between pseudo- and anonymization

One of the most relevant decisions on anonymization dates to 5 September 2024, namely the CEGEDIM SANTE case [23]. CEGEDIM designs and sells secretarial software for the medical sector. The company collected patient health data from doctors who agreed to participate in creating a health data repository. These data were allegedly anonymized with k-anonymity techniques. The decision was based on two key factors: (1) the WP29 test ¹⁴, which determines whether the data were anonymized, and (2) EUCJ case law. However, the rapporteur isolated a 6-year-old patient with a medical condition, which would suggest pseudonymization, unless re-identification is proved to be “negligible” by “reasonable means”. On this point, the CNIL concluded that the available data could be easily re-identified. However, the company contended that there were few educational materials on anonymization and that the guidelines lacked precision, rendering them unsuitable for legal certainty. Despite CNIL’s rejection of the complaint, the WP 29 Guidelines on anonymization are not updated concerning the current risks associated with k-anonymity. Furthermore, the CNIL has not specified what constitutes a ‘negligible risk’. This observation suggests that anonymity-washing may not be a deliberate practice.

¹⁴ Linkability, re-identification, and inference.

Take-away

Despite the consistency of the EUCJ case-law relating to personal data, confusion is widespread at the institutional level and among firms as far as the distinction between personal and non-personal data. Such confusion originates from the difficulty in distinguishing anonymization from pseudonymization techniques. This issue could enable unintentional anonymity-washing.

6 Discussion and overture

The gap between regulatory guidance and technical solutions gives space to anonymity-washing, whether intentional or unintentional. While frameworks exist to support anonymization efforts, their inconsistent application, misinterpretation, and continued reliance on outdated methods — repeatedly shown to be ineffective — often create a false sense of compliance and security. Several key issues contribute to this phenomenon.

In Section 3 we explain the **lack of clear guidance** to apply regulations and definitions. Guidelines lack a clear definition of pseudonymization and omit a definition of anonymization. This inconsistency can be a source of confusion and anonymity-washing. Nevertheless, establishing a coherent terminology is not a straightforward task, as personal data can be strongly situation-dependent, and data can be used in numerous ways.

We have also shown that **guidelines are often outdated or unreliable**. It has been shown that the anonymization techniques in the WP29 Opinion 5/2014 on anonymization are no longer reliable, and relevant questions persist, given the contradictory nature of the document. Consequently, guidelines on anonymization remain to be updated, and the information provided by the EDPB on pseudonymization and anonymization of the AI models does not address this gap.

Next, we have seen that the **differing interpretations** of the regulations among authorities compromise their uniform application, leading to legal uncertainty for businesses and organisations. For example, some authorities, like the CNIL, adopt a stringent approach, while others, like the ICO and the DPC, are more flexible.

Furthermore, in Section 5.1 we show that there is a **lack of awareness** among practitioners. They often do not recognize that their data can constitute personal data, leading to a failure to implement privacy by design principles [9]. Such misconceptions could be eliminated by adequate training and guidance that would give the right tools to practitioners and engineers to be able to competently assess their datasets and apply anonymization methods.

We believe that one crucial cause of this shortcoming is the **lack of understanding** of anonymization methods. Many practitioners, particularly those without advanced mathematical skills, struggle to understand and apply fundamental privacy principles. The complexity of privacy-enhancing technologies

(PETs) creates an additional barrier, making it difficult for non-specialists to implement effective anonymization (e.g.: [10,30,63]).

Another consequence of this educational deficiency is that organizations continue to **rely on outdated anonymization methods**, such as k-anonymity and l-diversity, despite their well-documented vulnerabilities [46]. This reliance stems from a lack of awareness regarding modern privacy-preserving techniques, as well as limited resources for evaluating and adopting alternative approaches.

However, this *lag* between the state-of-the-art and the most popular, but outdated tools is neither newfound nor unparalleled. There has always been a **collaboration gap** between academia and industry that limits the transfer of theoretical advancements into practice. Without structured mechanisms to facilitate knowledge-sharing, industry professionals may not only struggle to integrate the latest research into their anonymization strategies but also completely ignore it.

6.1 Overture

To address these issues and mitigate the risks of anonymity-washing, we would like to raise awareness among privacy experts and encourage them to facilitate adoption by practitioners for practitioners. With this objective in mind, we suggest the following actions:

We believe that one of the most important action to take is to **develop a comprehensive anonymization curriculum** that could be promoted and distributed by data protection authorities either in the form of training programs offered or thorough and up-to-date educational resources, such as books and hands-on exercises aided with structured guidance on fundamental privacy principles and their real-world applications. Key components should include: (1) A clear explanation of privacy threats and their manifestations in various datasets. (2) An overview of widely accepted privacy definitions (besides k-anonymity and l-diversity, adding differential privacy and cryptographic methods), including their advantages and drawbacks. (3) How these techniques can defend against said privacy threats. (4) Techniques for evaluating privacy technologies and applying them to real-world scenarios. (5) Strategies for auditing privacy risks and implementing mitigation measures in large-scale datasets. (6) Best practices for integrating privacy considerations into broader software engineering projects. (7) Case studies illustrating data breaches and the consequences of inadequate privacy protections. (8) Adding hands-on learning resources, such as Jupyter notebooks and real-world datasets.

Moreover, educational resources should be **tailored to diverse audiences**. Given the varying levels of expertise among practitioners, privacy education must be designed to accommodate both technical and non-technical professionals. It is also important to emphasize the **use of state-of-the-art methods**. Practitioners should be trained to critically assess privacy techniques and understand the limitations of traditional methods in modern, large datasets. In order to be sufficiently critical, we should also include adequate privacy risk assessments based on attacker capabilities, data sensitivity, and intended data

usage. Finally, we believe that a curriculum of this depth can not be delivered without **enhancing the collaboration between academia and industry**. We acknowledge that finding a common language between academia and industry is not always straightforward. It takes time and effort: joint initiatives, workshops, and training programs should be encouraged to bridge the gap between theoretical advancements and practical implementation. By realizing these recommendations, organizations and policymakers can move beyond superficial compliance efforts and work toward fostering a robust, meaningful approach to data privacy that will help reduce the prevalence of anonymity-washing and ensure that anonymization practices align with contemporary privacy risks, industrial demands, and capacities.

6.2 Future work

As a support of the collaboration between academia and industry, we wish to create a repository of guidelines and educational materials using the accumulated knowledge that we have used to write this paper. We envision constructing a proper website¹⁵ aided with instructions that would help practitioners navigate and find the appropriate guideline, document, or educational resource to their needs.

Furthermore, we conjecture that there is one more potential source of many of the aforementioned problems, namely, the use of popular anonymization tools. Thus, we have already started examining these existing tools; firstly, to corroborate our conjecture, and secondly, to be able to properly include them in the aforementioned website, equipping practitioners with the necessary understanding and comparison of these products.

References

1. Aggarwal, C.C.: On k-anonymity and the curse of dimensionality. In: VLDB. vol. 5, pp. 901–909 (2005)
2. Alik, E., Marietjie, B., Dusty-Lee, D., Beverley, T., Carmel, S., Donrich, T.: ‘potato potahto’? disentangling de-identification, anonymisation, and pseudonymisation for health research in africa. *Journal of Law and the Biosciences* **12**(1), 1 (2025)
3. Altman, M., Aloni Cohen, K.N.: *Acm techbriefs*. ACM TechBriefs (2024)
4. ANPD: Agenda regulatoria 2025-2026 (2025), <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-agenda-regulatoria-2025-2026>
5. ATSUMI, SAKAI: A guide to data protection in japan (2020), <https://www.aplawjapan.com/archives/pdf/data-protection-202009.pdf>
6. Ayyamperumal, S.G., Ge, L.: Current state of llm risks and ai guardrails. *arXiv preprint arXiv:2406.12934* (2024)
7. BfDI: Die anonymisierung im datenschutzrecht (2022), https://www.bfdi.bund.de/SharedDocs/Downloads/DE/DokumenteBfDI/Reden_Gastbeitr%C3%A4ge/2022/Anonymisierung-im-DS-recht.pdf?__blob=publicationFile&v=2

¹⁵ We plan to make a preliminary GitHub repo available upon acceptance of this manuscript.

8. BfDI: Aktuelle fragestellungen des datenschutzes (2023), https://www.bfdi.bund.de/SharedDocs/Downloads/DE/DokumenteBfDI/Reden_Gastbeitr%C3%A4ge/2023/eco-Kompetenzgruppe.pdf?__blob=publicationFile&v=2
9. BfDI: Arbeitspapier zu telemetrie und diagnosedaten (2023), https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Berlin-Group/20230608_AP-Telemetrie-Diagnosedaten.pdf?__blob=publicationFile&v=3
10. BfDI: Arbeitspapier zum thema "smart cities" (2023), https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Berlin-Group/20230608_WP-Smart-Cities.pdf?__blob=publicationFile&v=2
11. BfDI: Datenschutz durch technik – chancen und grenzen von anonymisierung, pseudonymisierung und pets (2024), https://www.bfdi.bund.de/SharedDocs/Downloads/DE/DokumenteBfDI/Reden_Gastbeitr%C3%A4ge/2024/Datenschutz-durch-Technik-BvD.pdf?__blob=publicationFile&v=1
12. BfDI: Datennutzung vs. datenschutz – veranstaltung zum europäischen datenschutztag (2025), https://www.bfdi.bund.de/SharedDocs/Downloads/DE/DokumenteBfDI/Reden_Gastbeitr%C3%A4ge/2025/Rede-Eu-Akademie-Informationsfreiheit-Datenschutz.pdf?__blob=publicationFile&v=2
13. Bhajaria, N.: Data Privacy: A runbook for engineers. Simon and Schuster (2022)
14. Brazil: Lei geral de proteção de dados pessoais (redação dada pela lei nº 13.853, de 2019) (lgpd) (2019)
15. Burkert, C., Federrath, H., Marx, M., Schwarz, M.: Positionspapier zur anonymisierung unter der dsgvo unter besonderer berücksichtigung der tk-branche. Konsultationsverfahren des BfDI **10** (2020)
16. Burt, A., Stalla-Bourdillon, S., Rossi, A.: A guide to the eu's unclear anonymization standards (2021), <https://iapp.org/news/a/a-guide-to-the-eus-unclear-anonymization-standards/>
17. CCPA: California consumer privacy act (ccpa) (2020)
18. Chevrier, R., Foufi, V., Gaudet-Blavignac, C., Robert, A., Lovis, C.: Use and understanding of anonymization and de-identification in the biomedical literature: scoping review. Journal of medical Internet research **21**(5), e13484 (2019)
19. Cirucci, A.M.: Oversharing the super safe stuff: "privacy-washing" in apple iphone and google pixel commercials. First Monday (2024)
20. CNIL: L'anonymisation des données, un traitement clé pour l'open data (2019), <https://www.cnil.fr/fr/lanonymisation-des-donnees-un-traitement-cle-pour-lopen-data>
21. CNIL: L'anonymisation de données personnelles (2020), <https://www.cnil.fr/fr/technologies/lanonymisation-de-donnees-personnelles>
22. CNIL: Recherche scientifique (hors santé) : enjeux et avantages de l'anonymisation et de la pseudonymisation (2022), <https://www.cnil.fr/fr/recherche-scientifique-hors-sante-enjeux-et-avantages-de-lanonymisation-et-de-la-pseudonymisation>
23. nationale de l'informatique et des libertés CNIL, C.: Délibération san-2024-013 (5 September 2024)
24. Cohen, A., Nissim, K.: Towards formalizing the gdpr's notion of singling out. Proceedings of the National Academy of Sciences **117**(15), 8344–8352 (2020)
25. Commission, F.T.: Protecting consumer privacy in an era of rapid change, recommendations for businesses and policymakers, ftc report (2012), <https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf>

26. Cormode, G., Srivastava, D., Li, N., Li, T.: Minimizing minimality and maximizing utility: analyzing method-based attacks on anonymized data. *Proceedings of the VLDB Endowment* **3**(1-2), 1045–1056 (2010)
27. Craig, T., Ludloff, M.E.: Privacy and big data: the players, regulators, and stakeholders. " O'Reilly Media, Inc." (2011)
28. De Montjoye, Y.A., Hidalgo, C.A., Verleysen, M., Blondel, V.D.: Unique in the crowd: The privacy bounds of human mobility. *Scientific reports* **3**(1), 1376 (2013)
29. De Montjoye, Y.A., Pentland, A.S.: Response to comment on "unique in the shopping mall: On the reidentifiability of credit card metadata". *Science* **351**(6279), 1274–1274 (2016)
30. Djiriguian, Missègue, R.: Diffuser une base anonymisée : utopie ou réalité? *Journées de méthodologie statistique de l'Insee* (2022), <https://journées-méthodologie-statistique.insee.net/gestion-du-secret-pour-la-diffusion-grand-public-de-cubes-multidimensionnels-une-experimentation-au-ssm-agriculture/>
31. DPC: Data protection: The basics (2019), <https://dataprotection.ie/sites/default/files/uploads/2019-07/190710%20Data%20Protection%20Basics.pdf>
32. DPC: Guidance on anonymisation and pseudonymisation (2019), <https://www.dataprotection.ie/sites/default/files/uploads/2022-04/Anonymisation%20and%20Pseudonymisation%20-%20latest%20April%202022.pdf>
33. EC, E.C.: Directive 95/46/ec of the european parliament and of the council on the protection of individuals with regard to the processing of personal data and on the free movement of such data, ojl 281 (1995)
34. EC, E.C.: Communication from the commission to the european parliament and the council, guidance on the regulation on a framework for the free flow of non-personal data in the european union, com/2019/250 final (2019)
35. EC, E.C.: Communication from the commission to the european parliament and the council - two years of application of the general data protection regulation, com/2020/264 (2020)
36. EC, E.C.: Communication from the commission to the european parliament and the council - second report on the application of the general data protection regulation, com/2024/357 (2024)
37. EDPB, A.: 10 misunderstandings related to anonymisation (2021)
38. EDPB, E.D.P.B.: Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of ai models (2024)
39. EDPB, E.D.P.B.: Guidelines on pseudonymisation - version for public consultation (2025)
40. El Emam, K., Alvarez, C.: 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)
41. Elliot, M., Mackey, E., O'Hara, K.: The anonymisation decision-making framework 2nd edition: European practitioners' guide (2020)
42. EU-USA: Privacy shield framework (2016), https://eur-lex.europa.eu/eli/dec_impl/2016/1250/oj/eng
43. EUGC, E.U.G.C.: Srb vs edps, case t-557/20 (26 April 2023)
44. European Parliament, Council of the European Union: Regulation (EU) 2016/679 of the European Parliament and of the Council (gdpr) (2016), <https://data.europa.eu/eli/reg/2016/679/oj>
45. European Parliament, Directorate-General for Internal Policies, P.D.E., Policy, S.: Industry, research and energy, data flows- future scenarios, in-depth analysis for the itre committee (2017)

46. Gadotti, A., Rocher, L., Houssiau, F., Crețu, A.M., De Montjoye, Y.A.: Anonymization: The imperfect science of using data while preserving privacy. *Science Advances* **10**(29), eadn7053 (2024)
47. Garfinkel, S.L.: De-identification of personal information. National Institute of Standard and Technology (2015)
48. GPDP: Provvedimento [10090499] (13 November 2024)
49. GPDP: Guida all'applicazione del regolamento europeo in materia di protezione dei dati personali (2023), <https://www.garanteprivacy.it/documents/10160/0/Guida+all+applicazione+del+Regolamento+UE+2016+679.pdf/2281f960-a7b2-4c53-a3f1-ad7578f8761d?version=2.0>
50. ICO: Anonymisation: managing data protection risk code of practice (2019)
51. ICO: (draft) anonymisation, pseudonymisation and privacy enhancing technologies guidance (2022), <https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-call-for-views-anonymisation-pseudonymisation-and-privacy-enhancing-technologies-guidance/>
52. ICO: Privacy-enhancing technologies (pets) (2023), <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/privacy-enhancing-technologies/>
53. of India, D.S.C.: Balancing privacy and innovation: Anonymisation standards for indian data (2023)
54. INSEE: Risque de ré-identification : deux questions pratiques relatives au critère de la l-diversité (2019), <https://www.insee.fr/fr/information/4277545>
55. INSEE: Guide du secret statistique (2024), <https://www.insee.fr/fr/information/1300624>
56. INSEE: Les méthodes perturbatives d'anonymisation des données individuelles (2024), https://www.insee.fr/fr/statistiques/fichier/4277545/1-SMS_secret_24_juin_2019.pdf
57. INSEE, M.B.: La question de la confidentialité des données individuelles (2016), <https://www.insee.fr/fr/statistiques/2535625>
58. Japan: Act on the protection of personal information (act no. 57 of 2003) (last version act no. 37 of 2021) (2003), <https://www.japaneselawtranslation.go.jp/en/laws/view/4241/en>
59. Jarmul, K.: Practical data privacy. " O'Reilly Media, Inc." (2023)
60. of Justice EUCJ, E.U.C.: Judgment of the court (grand chamber): Data protection commissioner v facebook ireland limited and maximillian schrems (16 July 2020)
61. of Justice EUCJ, E.U.C.: Breyer, case c-582/14 (19 October 2016)
62. of Justice EUCJ, E.U.C.: Goc vs european commission, case c-479/22 p. (7 March 2024)
63. of Justice EUCJ, E.U.C.: Iab europe, case c-604/22. (7 March 2024)
64. of Justice EUCJ, E.U.C.: Gesamtverband autoteile-handel (accès aux informations sur les véhicules), case c-319/22 (9 Novembre 2023)
65. Langarizadeh, M., Orooji, A., Sheikhtaheri, A.: Effectiveness of anonymization methods in preserving patients' privacy: a systematic literature review. *Health Informatics Meets eHealth* pp. 80–87 (2018)
66. Levi-Valensin, M.: Gestion du secret pour la diffusion grand public de cubes multi-dimensionnels: une expérimentation au ssm agriculture. *Journées de méthodologie statistique de l'Insee* (2022)
67. Mamanazarov, S.: De-identification and anonymization: legal and technical approaches (2024)

68. Moon-Ho Joo, H.Y.K.: Comparison of personal information de-identification policies and laws within the eu, the us, japan, and south korea. *Government Information Quarterly* **40**, 101805 (2023)
69. Narayanan, A., Felten, E.W.: No silver bullet: De-identification still doesn't work. *White Paper* **8** (2014)
70. Narayanan, A., Shmatikov, V.: Robust de-anonymization of large sparse datasets: a decade later. May **21**, 2019 (2019)
71. Nguyen, B.: Techniques d'anonymisation. *Statistique et société* **2**(4), 53–60 (2014)
72. Nguyen, B., Castelluccia, C.: Techniques d'anonymisation tabulaire: concepts et mise en oeuvre. *arXiv preprint arXiv:2001.02650* (2020)
73. Nissenbaum, H.: Privacy in context: Technology, policy, and the integrity of social life. In: *Privacy in context*. Stanford University Press (2009)
74. OECD: Cross-border data flows, policy sub-issue (2022), <https://www.oecd.org/en/topics/cross-border-data-flows.html>
75. Ohm, P.: Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA L. Rev.* **57**, 1701 (2009)
76. Parliament, E., of the Council: Regulation (eu) 2024/1689 of the european parliament and of the council of 13 june 2024 laying down harmonised rules on artificial intelligence and amending regulations (ec) no 300/2008, (eu) no 167/2013, (eu) no 168/2013, (eu) 2018/858, (eu) 2018/1139 and (eu) 2019/2144 and directives 2014/90/eu, (eu) 2016/797 and (eu) 2020/1828 (artificial intelligence act), pe/24/2024/rev/1, oj l, 2024/1689, (2024)
77. Parliament, E., of the Council: Regulation (eu) 2018/1725 on the protection of natural persons with regard to the processing of personal data by the union institutions, bodies, offices and agencies and on the free movement of such data, and repealing regulation (ec) no 45/2001 and decision no 1247/2002/ec, pe/31/2018/rev/1, oj l 295, 21.11.2018, p. 39–98, art 15 (1) d (23 October 2018)
78. Party, A..D.P.W.: Opinion 05/2014 on anonymisation techniques (2014)
79. Party, A..D.P.W.: Opinion on the concept of personal data (2007)
80. PDPC: Guide to basic anonymisation (2022)
81. Peloquin, D., DiMaio, M., Bierer, B., Barnes, M.: Disruptive and avoidable: Gdpr challenges to secondary research uses of data. *European Journal of Human Genetics* **28**(6), 697–705 (2020)
82. Québec: Regulation respecting the anonymization of personal information (2024)
83. Rubinstein, I.S., Hartzog, W.: Anonymization and risk. *Wash. L. Rev.* **91**, 703 (2016)
84. council of Advisors on Science, P., Technology, W.H.: Big data and privacy: A technological perspective (2014)
85. Secretariat, P.I.P.C.: Anonymously processed information - towards balanced promotion of personal data utilization and consumer trust (2017), https://www.ppc.go.jp/files/pdf/The_PPC_Secretariat_Report_on_Anonymously_Processed_Information.pdf
86. Sharma, S.: Data privacy and GDPR handbook. John Wiley & Sons (2019)
87. Stadler, T., Troncoso, C.: Why the search for a privacy-preserving data sharing mechanism is failing. *Nature Computational Science* **2**(4), 208–210 (2022)
88. Stalla-Bourdillon, S., Burt, A.: The definition of 'anonymization' is changing in the eu: Here's what that means (2023), <https://iapp.org/news/a/the-definition-of-anonymization-is-changing-in-the-eu-heres-what-that-means>
89. Stalla-Bourdillon, S.: Identifiability, as a data risk: Is a uniform approach to anonymisation about to emerge in the eu? Available at SSRN (2025)

90. Stalla-Bourdillon, S., Knight, A.: Anonymous data v. personal data-false debate: an eu perspective on anonymization, pseudonymization and personal data. *Wis. Int'l LJ* **34**, 284 (2016)
91. Stallings, W.: Information privacy engineering and privacy by design: Understanding privacy threats, technology, and regulations based on standards and best practices. Addison-Wesley Professional (2019)
92. Sénéchal, J.: Publication de l'avis de l'edpb du 17 décembre 2024 sur le traitement des données personnelles dans le contexte des modèles d'ia : prémices d'une mutation profonde du rgpd ? (2024), <https://www.dalloz-actualite.fr/flash/publication-de-l-avis-de-l-edpb-du-17-decembre-2024-sur-traitement-des-donnees-personnelles-da>
93. USA: Health insurance portability and accountability act (1996)
94. Wallace, S.E.: What does anonymization mean? datashield and the need for consensus on anonymization terminology. *Biopreservation and Biobanking* **14**(3), 224 (2016)
95. Weitzenboeck, E.M., Lison, P., Cyndecka, M., Langford, M.: The gdpr and unstructured data: is anonymization possible? *International Data Privacy Law* **12**(3), 184–206 (2022)
96. Wolff, J., Lehr, W., Yoo, C.S.: Lessons from gdpr for ai policymaking. *Va. JL & Tech.* **27**, 1 (2023)
97. Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., Nissim, K., O'Brien, D.R., Steinke, T., Vadhan, S.: Differential privacy: A primer for a non-technical audience. *Vand. J. Ent. & Tech. L.* **21**, 209 (2018)