Anonymization did *not* fail: Misconceptions and overstatements on data anonymization failures

David Sánchez, *Universitat Rovira i Virgili, CYBERCAT-Center for Cybersecurity Research of Catalonia, Tarragona, 43007, Catalonia*

Josep Domingo-Ferrer, *Universitat Rovira i Virgili, CYBERCAT-Center for Cybersecurity Research of Catalonia, Tarragona, 43007, Catalonia, and LAAS-CNRS, Université de Toulouse, 7 Av. du Colonel Roche, Toulouse, 31400, France*

Krishnamurty Muralidhar, *University of Oklahoma, Norman, OK 73019, U.S.A.*

Abstract—Research on data anonymization has been carried out for more than 50 years. Yet, several authors have claimed the "failure of anonymization". We review the most publicized privacy leaks reported over the past decades and conclude that they were due to non-existent or inadequate anonymization, rather than to a lack of robust anonymization methods.

Introduction

Research on data anonymization has been carried out since the 1960s with the goal of publishing analytically useful data that cannot be linked to specific individuals or reveal confidential information on them. As a result, there is currently a wide choice of anonymization methods in the literature for both statistical disclosure control [9] and privacy-preserving data publishing [7]. However, from time to time and with increasing frequency, a new and highly publicized headline appears announcing "the failure of anonymization" [1], [3], [4], [5], [10], [11], [13], [16], [17]. These headlines have a serious impact on society and policy making. For example, President Biden's Executive Order 14177 of Feb. 28, 2024 "Preventing Access to Americans' Bulk Sensitive Personal Data and U.S. Government-Related Data by Countries of Concern" stated that de-anonymizing is increasingly feasible due to recent technical advances and easier data access [a].

Due to the adverse consequences that such claims may have on the willingness of data owners to share the data so necessary for research, and their influence on the definition of too restrictive data regulations, we aim to shed light on the alleged ineffectiveness of anonymization methods, most of them available for several decades.

XXXX-XXX © 2025 IEEE
Digital Object Identifier 10.1109/XXX.0000.0000000

[a] https://www.federalregister.gov/documents/2024/03/01/2024-04573/preventing-access-to-americans-bulk-sensitive-personal-data-and-united-states-government-related

Fundamentals of data anonymization

To understand and analyze the roots of these claims, we first need to recall the fundamental notions regarding disclosure and privacy protection.

There are two types of disclosure:

- *Identity disclosure* allows associating some published unidentified record with the subject to whom it corresponds. This process is also known as *re-identification*.
- *Attribute disclosure* enables determining the value of a subject's confidential attribute, such as income or diagnosis. Attribute disclosure can occur with or without identity disclosure.

Anonymization focuses on preventing re-identification by altering the published information so that the data subject can no longer be identified directly or indirectly. Attributes that enable direct re-identification are *personal identifiers* (such as passport numbers), whereas *quasi-identifiers* (*e.g.*, zip code, gender or age) are those that do not uniquely identify the subject, but whose combination may because they may be present in public identified databases like electoral rolls. Proper anonymization, as it has been understood in the literature for nearly half a century,

must *remove identifiers and mask all quasi-identifiers* to truly prevent re-identification. In other words, any procedure that does not consider *both* identifiers and quasi-identifiers should not be called anonymization, because it does not preclude identity disclosure. Examples of deficient privacy protection procedures are

- *De-identification*, which only removes direct identifiers or, at most, a limited number of predefined quasi-identifiers.
- *Pseudonymization*, which replaces direct identifiers with artificial identifiers or pseudonyms. Not only is pseudonymization insufficient to avoid re-identification but, unlike anonymization and de-identification, pseudonymization is also reversible and allows linkage among different records referring to the same subject, thereby increasing the risk of disclosure. In fact, pseudonymized data still qualify as *personal data* under the European General Data Protection Regulation (GDPR) and therefore cannot be freely shared without additional safeguards [8].

In contrast, the above-mentioned sound anonymization masking —backed by *ex post* disclosure risk assessment— has been in use for almost 50 years. Furthermore, privacy models giving formal anonymization guarantees against re-identification have been around for more than 25 years (the k-anonymity privacy model), and privacy models preventing attribute disclosure have existed for about 20 years (*e.g.*, k-anonymity extensions such as l-diversity and t-closeness) [7].

Examining the alleged failures of anonymization

Unfortunately, the terms anonymization, de-identification, and pseudonymization are all too often used almost interchangeably. This happens in the media and even in some scientific papers, as is apparent in the most notorious cases of "anonymization failures":

- In 1997, the Governor of Massachusetts was re-identified in his *de-identified* medical records (where only obvious identifiers such as name, address, and Social Security numbers had been removed) by using publicly accessible voting rolls [16].
- In 2006, The New York Times successfully uncovered the identities of several users in the AOL Web search query logs, which were released by the company after only *pseudonymizing* user identities through numerically identified accounts [1]. Several real identities were easily obtained by cross-referencing search records with publicly available identified phonebook listings.
- In 2007, [10] reported that they had obtained the usernames of several individual users from the Netflix Prize data set (consisting of *pseudonymized* movie ratings by Netflix users) by matching the data with username-identified film ratings on the Internet Movie Database.
- In 2014, a researcher re-identified the taxi rides of several Hollywood celebrities in the NYC Taxicab data set, in which the taxi identifiers were *pseudonymized* via easily breakable hashing [17].

It is obvious that none of the above data sets were *anonymized* and, in most cases, re-identification happened by matching non-masked quasi-identifiers with publicly available identified sources, which is *precisely what proper anonymization methods prevent*. However, most of the aforementioned works, the media and professionals in different disciplines considered these cases as "anonymization failures" and questioned the effectiveness of anonymization methods [11].

Even for poorly protected data sets, re-identification is not as straightforward as sometimes suggested. For re-identification to occur, it must be unequivocal, and for this, the following four conditions must be met:

1) There must be records in the published data set with unique combinations of quasi-identifying values.
2) The published data set should be an exhaustive representation of a population. Otherwise, uniqueness in the released data set does not necessarily imply uniqueness in the population.
3) Knowledge of the quasi-identifiers used in the attack should be plausible. Assuming knowledge of too specific or too much information on the subject makes re-identification contrived and implausible.
4) If the quasi-identifiers used in the attack are inferred (*e.g.*, through machine learning models) rather than known, the inference must be 100% accurate and must not generate false positives.

In this respect, a series of papers rediscovered the notion of record uniqueness and its role in re-identification (a topic present in the statistical disclosure control literature since the 1970s) and reported extremely large re-identification figures even in "anonymized" data sets [3], [4], [13]. Specifically, in

[3] it is stated that "four spatio-temporal points are enough to uniquely identity 95% of the individuals" in a *de-identified* cell-phone data set, whereas in [4] it is claimed that 90% individuals can be re-identified in a *de-identified* data set of credit card transactions; finally, in [13] it is claimed that "99.98% of Americans would be correctly re-identified in any data set using 15 demographic attributes" by employing generative models. However, *not one reidentification is actually reported*.

On the one hand, the data sets used were not actually anonymized, as discussed in [15]. On the other hand, the re-identification claims were overstated, as the aforementioned 2nd, 3rd and 4th conditions needed for unequivocal re-identification to happen were not met. In fact, the authors mistook *re-identification risk* for actual *re-identification*. They went on to implement some *ad hoc* "anonymization" strategies to coarsen data that proved ineffective to reduce uniqueness; and, from this, they concluded that anonymization methods in general are ineffective [4]. Even though data coarsening is indeed a method often used in anonymization to mask quasi-identifiers, the *ad hoc* coarsening approach used was flawed, as discussed in [15]: quasi-identifiers were independently coarsened using value ranges fixed *ex ante*, which is inappropriate for at least two reasons: (i) to offer true anonymity guarantees, coarsening should be based on the actual distribution of the data set; (ii) independently coarsening each quasi-identifier cannot ensure that unique value combinations disappear — this requires coarsening *all* quasi-identifiers together, as in *k*-anonymity.

A recent trend is to challenge anonymization using *database reconstruction attacks*. Given some released (protected) statistical outputs, the objective of a reconstruction attack is to come up with a candidate original data set that is compatible with the released outputs. The alleged disclosure risks resulting from reconstruction attacks were so publicized [b] that the U.S. Census Bureau decided to replace the anonymization methodology they had been employing for Decennial Censuses up to 2010 (and for which no single re-identification was ever reported) with a method based on differential privacy. Although differential privacy is a sound privacy model, it is hard to enforce in practice due to the significant logical inconsistencies and inaccuracies introduced by the noise addition it requires. This was the case for the release of the 2020 Decennial Census, whose data were not accurate enough for a substantial range of analyses [c].

In reality, database reconstruction attacks should be a concern only if they can *uniquely* link a record of the (unidentified) reconstructed data to a record in the (identified) attacker's data. However, this cannot occur if there are several reconstructions compatible with the released outputs, which is the case for all the reconstructions reported so far [14]. For example, for the recent attempt to reconstruct Census data in [5], the number of reconstructed records was up to 48 times greater than the number of original records [14]. Such a large degree of uncertainty in the reconstruction poses no real privacy threat.

Privacy risks in AI

The advent of modern AI models has reignited the debate on privacy risks and the effectiveness of data anonymization. Fears of data disclosure in AI models have materialized in restrictive regulations, such as Biden's Executive Order 14110 of Oct. 30, 2023 "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" or the recent EU Artificial Intelligence Act [d]. Given that EO 14110 was revoked by Trump to increase the freedom of the US AI industry, special care should be taken in other jurisdictions to precisely gauge AI privacy risks and make sure that any legal constraints are really justified by the state of the technology.

The most widespread privacy concern on AI models is their vulnerability to membership inference attacks (MIAs), which are the basis for more sophisticated threats such as data reconstruction attacks. MIAs aim to determine whether a specific data point was included in a model's training data by exploiting differences in model behavior when presented with member and non-member data points.

However, inferring membership in training data does not pose a privacy threat in itself. In fact, it only poses a threat if confirming membership unequivocally exposes some confidential information (*i.e.*, attribute disclosure), such as when a model is trained on medical records of patients with a *common* sensitive condition. Although this is a scenario frequently mentioned in the AI literature as a motivating case, it is quite rare in practice, especially when training general AI models

[b] https://www2.census.gov/about/partners/cac/sac/meetings/2022-09/presentation-reconstruction-and-re-dentification-of-dhc-file.pdf

[c] https://www2.census.gov/library/publications/decennial/2020/2020-census-disclosure-avoidance-handbook.pdf

[d] https://artificialintelligenceact.eu/

such as foundational models. In addition, the inference performance of MIAs is very low when evaluated under realistic conditions, such as well-generalizable target models [6]. It turns out that most MIAs in the AI literature are mounted against overfitted models to show better performance, even if overfitting is undesirable in real-world production models. Finally, similar to what was discussed above for re-identification disclosure, for membership disclosure to be a real privacy threat, it should be unequivocal. This is very rare in MIAs, as they tend to exhibit high false positive rates due to the frequent misclassification of non-members that are semantically similar to members [12].

Despite the poor performance of MIAs in realistic settings (even against completely unprotected models), if countering such attacks is necessary, it can be effectively done by general-purpose data protection mechanisms. Examples of defenses are differential privacy or anti-overfitting mechanisms [2]; the latter are more targeted, as overfitting is the crux of disclosure in AI models.

Generative AI poses new threats that are currently being investigated due to its ability to memorize and regenerate training data. However, generative AI models cannot reveal more information than was used to train them. If robust data anonymization is implemented on the training data (as should be done and has been done in statistical data releases for years), disclosure in generative AI will be under control.

Thus, the privacy risks inherent in AI can be effectively mitigated with the help of several techniques, including anonymization in a prominent place.

Anonymization still holds

We have shown that (i) privacy leaks reported over the past 30 years are due to nonexistent or inadequate anonymization, rather than to a lack of robust anonymization methods; (ii) claims on the ineffectiveness of anonymization methods come from misunderstandings that are attributable to a lack of familiarity with the anonymization literature that has developed over nearly five decades; (iii) privacy threats in AI are weak, especially in realistic and competitive AI models, and can be effectively mitigated via existing data protection mechanisms.

Unfortunately, maybe because sensational claims sell better, the few papers on attacks that we have discussed in the third section have attracted more than 10,000 citations from other scientific articles[e]. This has caused significant and undeserved damage not only to the data anonymization field but to research in general, as such unfounded alarms may dissuade or prevent the release of anonymized open data or the participation of the public in important studies that are much needed for the progress of science, transparency, and economic growth.

And yet, as of today, even with the enormous computational capabilities and sophisticated machine learning models available, plus the easy access to identified sources of personal data (*e.g.*, social networks), to the best of our knowledge, *no single successful re-identification has been reported on properly anonymized data* —such as those released by national statistical institutes, where identifiers are removed and quasi-identifiers are masked at record level.

Given this state of affairs, data owners and subjects can be reassured that there are sound anonymization methods to produce anonymized data that can be and are being safely shared. Regulators should also adopt a more informed (rather than preventive) approach to avoid imposing disproportionate burdens, especially on AI providers and deployers, which might unnecessarily compromise the competitiveness of the industry.

[e]Google Scholar (June, 2025)

ACKNOWLEDGMENTS

Partial support for this work has been received from the Government of Catalonia (ICREA Acadèmia Prizes to J. Domingo-Ferrer and to D. Sánchez, and grant 2021SGR-00115), MCIN/AEI/10.13039/501100011033 and "ERDF A way of making Europe" under grant PID2021-123637NB-I00 "CURLING", the EU's NextGenerationEU/PRTR via INCIBE (project "HERMES" and INCIBE-URV cybersecurity chair), and the Centre International de Mathématiques et d'Informatique de Toulouse.

REFERENCES

1. Barbaro, M., Zeller, T. A face is exposed for AOL searcher no. 4417749. *The New York Times* (Aug. 9, 2006); https://www.nytimes.com/2006/08/09/technology/09aol.html
2. Blanco-Justicia, A., Sánchez, D., Domingo-Ferrer,J., Muralidhar, K. A critical review on the use (and misuse) of differential privacy in machine learning. *ACM Computing Surveys* 55 (2023) 160:1-160:16.
3. 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*, 3 (2013), 1376.

4. de Montjoye, Y.-A., Radaelli, L., Singh, V.K., Pentland, A.S. Unique in the shopping mall: on the reidentifiability of credit card metadata. *Science 347*, 6221 (2015), 536–539.
5. Dick, T., Dwork, C., Kearns, M., Liu, T., Roth, A., Vietri, G., Wu, Z.S. Confidence-ranked reconstruction of census microdata from published statistics. *Proceedings of the National Academy of Sciences U.S.A. 120*, 8 (2023), e2218605120.
6. Dionysiou, A., Athanasopoulos E. Sok: Membership inference is harder than previously thought. *Privacy Enhancing Technologies Symposium* (2023).
7. Domingo-Ferrer, J., Sánchez, D., Soria-Comas, J. *Database anonymization: privacy models, data utility and microaggregation-based inter-model connections*. Synthesis Lectures on Information Security and Privacy. Morgan & Claypool Publishers (2016).
8. *Guidelines 01/2025 on Pseudonymization*, European Data Protection Board, 2025. https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_en
9. Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K., De Wolf, P.-P. *Statistical Disclosure Control*. Wiley (2012).
10. Narayanan, A., Shmatikov, V. Robust deanonymization of large sparse data sets. *Proceedings of the IEEE Symposium on Security and Privacy* (2008), 111–125.
11. Ohm, P. Broken promises of privacy: responding to the surprising failure of anonymization. *UCLA Law Review 57* (2009), 1701.
12. Rezaei, S., Liu, X. On the difficulty of membership inference attacks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), 7892–7900.
13. Rocher, L., Hendrickx, J.M., de Montjoye, Y.-A. Estimating the success of re-identifications in incomplete data sets using generative models. *Nature Communications 10*, 1 (2019), 1–9.
14. Sánchez, D., Jebreel, N., Muralidhar, K., Domingo-Ferrer, J., Blanco-Justicia, A. An examination of the alleged privacy threats of confidence-ranked reconstruction of census microdata. *Proceedings of Privacy in Statistical Databases* (2024), 213–224.
15. Sánchez, D., Martínez, S., Domingo-Ferrer, J. Comment on 'Unique in the shopping mall: On the reidentifiability of credit card metadata'. *Science 351*, 6279 (2016), 1274.
16. Sweeney, L. Simple demographics often identify people uniquely. *Health 671* (2000), 1–34.
17. Tockar, A. Riding with the stars: passenger privacy in the NYC Taxicab data set (Sep. 15, 2014); https://agkn.wordpress.com/author/atockar/

David Sánchez (Senior Member, IEEE) received a PhD in computer science from the Technical University of Catalonia in 2008. He is a full professor and an ICREA-Acadèmia research professor at Universitat Rovira i Virgili, Tarragona, Catalonia. His research interests include data semantics, machine learning, and data privacy. Contact him at david.sanchez@urv.cat.

Josep Domingo-Ferrer (Fellow, IEEE) received BSc-MSc and PhD degrees in computer science (Autonomous University of Barcelona), a BSc-MSc in mathematics (UNED) and an MA in philosophy (U. Paris Nanterre). He is a distinguished full professor of computer science and an ICREA-Acadèmia research professor at Universitat Rovira i Virgili, Tarragona, Catalonia, where he also leads CYBERCAT. He currently is an invited professor at LAAS-CNRS, Université de Toulouse. His research interests include data privacy, data security, and ethics in IT. Contact him at josep.domingo@urv.cat

Krishnamurty Muralidhar received his undergraduate degree (1977) from the University of Madras, his graduate degree (1982) from Sam Houston State University, and his PhD (1986) from Texas A&M University. He is a full professor at the University of Oklahoma. His main research interest is data privacy. His research is on developing techniques that allow organizations to securely analyze, share, and disseminate confidential data without compromising privacy. Dr. Muralidhar received a patent for Data shuffling – a procedure that allows organizations to securely release or share sensitive data. Contact him at krishm@ou.edu