The lawfulness of re-identification under data protection law

Teodora Curelariu, Alexandre Lodie

The lawfulness of re-identification under data protection law

Teodora Curelariu[1] and Alexandre Lodie[2]

[1] CESICE, Université Grenoble-Alpes, Centre Inria de l'Université Grenoble-Alpes, France
[2] Centre Inria de l'Université Grenoble-Alpes, France

Abstract. Data re-identification methods are becoming increasingly sophisticated and can lead to disastrous data breaches. Re-identification is a key research topic for computer scientists as it can be used to reveal vulnerabilities of de-identification methods such as anonymisation or pseudonymisation. However, re-identification, even for research purposes, involves processing personal data. From this background, this paper aims to investigate whether re-identification carried out by computer scientists for research purposes can be considered GDPR-compliant. This issue is paramount to contribute to improving the state of knowledge concerning data security measures.

Keywords: Re-identification, Computer Science, PETs, Personal Data, GDPR, Data Protection

1 Introduction

Data are being increasingly shared on a wide scale, be it for public re-use[1] but also for marketing purposes.[23] Nonetheless, this trend comes with significant privacy concerns,[4] particularly regarding the re-identification of individuals through data mining.[5]
This poses risks such as unauthorised access, misuse of personal information and disclosure of personal data, ultimately undermining de-identification techniques and privacy safeguards.

De-identification, as defined by the NIST,[6] serves as a mechanism for organisations "to remove personal information from data that they collect, use, archive, and share with other organisations.[7]" As for re-identification, it can be defined as "a process by which information is

1 Peloquin, D., DiMaio, M., Bierer, B. et al. Disruptive and avoidable: GDPR challenges to secondary research uses of data. Eur J Hum Genet 28, 697–705 (2020).
2 Sheth, J., Charles H., Next Frontiers of Research in Data Driven Marketing: Will Technique Keep up with Data Tsunami?. Journal of Business Research, vol. 125, 780–84 (2021).
3 Rogers, J., Song A., Digital Marketing in The Legal Profession: What's Going On and Does It Matter? Law, Technology and Humans, vol. 5, no. 2, 134 – 64 (2023).
4 Henriksen-Bulmer, J., Jeary, S.: Re-identification attacks—A systematic literature re-view, International Journal of Information Management, 36, 1184–1192 (2016).
5 Schermer, B.W., The Limits of Privacy in Automated Profiling and Data Mining. Computer Law & Security Review, (2011).
6 National Institute of Standards and Technology
7 NIST, https://www.nist.gov/itl/iad/deidentificationnistgov, last accessed 2024/02/22.

attributed to de-identified data in order to identify the individual to whom the de-identified data relate.[8]"

Re-identification of de-identified data challenges the effectiveness of data protection techniques, which are designed to safeguard personal data by preventing the direct or indirect identification of data subjects. From this standpoint, re-identification is perceived as a security risk that must be dealt with by computer scientists and legal practitioners. The very definition of personal data underscores that "personal data are any information which are related to an identified or identifiable natural person.[9]" It is worth mentioning that according to these definitions, identifiability is the core criterion to define personal data. The importance of identifiability has been underlined by the literature as well.[10] Despite its importance, the definition of identifiability remains uncertain. Indeed, this is a contentious issue which is widely debated by scholars, some of them advocating for an objective approach of the identifiability of data subjects, while other stand for a more relative approach.[11] As regards the former approach, the qualification of data as personal data depends on the inherent features of the data themselves.[12] The focus is on whether data alone permit the re-identification of data subjects, no matter who holds them. When it comes to the latter approach, identifiability is more context-related and depends on the means and additional information in the hands of the person or organisation holding data.[13]

Thus, the choice of approach can directly influence the way data protection law applies. When the objective approach prevails, data may be considered personal if there is an abstract possibility to re-identify them. This broad interpretation places a greater burden on data controllers to ensure compliance with privacy regulations and to implement robust measures to personal data.

A contrario, in jurisdictions favouring the relative approach, the determination of whether data qualifies as personal may depend on the specific circumstances surrounding its processing and the likelihood of re-identification. This approach offers more flexibility, for instance in matter of data re-use for research purposes,[14] but may also create uncertainty and inconsistency in legal interpretations.[15] To sum up, the identifiability criterion involves assessing the likelihood of re-identification of data subjects,[16] and thus the applicability of data protection regulations.

To assess the likelihood of re-identification, a data controller must take into account some objective factors such as "the available technology at the time of the processing and technologi-

---

8 NIST, https://csrc.nist.gov/glossary/term/re_identification, last accessed 2024/02/22.

9 See Article 4 of the GDPR.

10 Spindler, G., Schmechel, P.: Personal Data and Encryption in the European General Data Protection Regulation., JIPITEC, 164 (2016).

11 Zuiderveen Borgesius, F., The Breyer Case of the Court of Justice of the European Union: IP Addresses and the Personal Data Definition., European Data Protection Law Review, Vol 3, Issue 1, 130–137 (2017).

12 OPINION OF ADVOCATE GENERAL CAMPOS SÁNCHEZ-BORDONA delivered on 12 May 2016, Case C-582/14 Patrick Breyer v Bundesrepublik Deutschland, § 52.

13 Ibid., § 53.

14 Mourby, M. et al.: Are 'pseudonymised' data always personal data? Implications of the GDP for administrative data research in the UK., Computer Law & Security Review, Vol 34, 222 – 233, (2018).

15 Lodie A., Case C-479/22 P, Case C-604/22 and the limitation of the relative approach of the definition of 'personal data' by the ECJ., European Law Blog, (2024).

16 See recital 26 of the GDPR.

cal developments.[17]". Put differently, the development of re-identification techniques and technology has an impact on the assessment of the robustness of a de-identification scheme. It contributes to what is considered the "state-of-the-art" in matter of data security.[18] The more sophisticated re-identification is, the stronger de-identification will be. Once de-identification is considered to be achieved and involves only a residual risk of re-identification, data are said to be anonymised.[19]

Quite surprisingly, re-identification is not considered as an independent research area, at least for legal scholars. It is mainly seen as a means to assess the reliability of data protection techniques and whether data have been properly anonymised with regard to the requirements of the GDPR.[20]

De-identification involves various data protection methods including anonymisation, but also pseudonymisation. The latter is mentioned as a security measure which can help protect data under Article 32 of the GDPR. This article underlines that data must be granted an appropriate level of security, taking into account the "state-of-the-art". These elements show that there is a pressing need for computer science research on this field. No matter if data are said to be anonymised or only pseudonymised, they can be re-identified so that re-identification is a risk to be taken into account when implementing data protection techniques.

As data controllers have obligations regarding the implementation of cutting-edge security measures, research is needed in order to keep de-identification techniques up-to-date. However, the lawfulness of re-identification under EU data protection law remains uncertain, as it inherently involves bypassing data protection measures.

This paper aims to investigate why carrying out research in the field of re-identification is key to improve privacy. It also intends to assess whether re-identification techniques implemented for research purposes can be considered GDPR compliant.

In the following section, technical aspects related to re-identification are discussed, and how research in this area contributes to enhancing privacy. Building on this technical foundation, Section 3 will question the lawfulness of re-identification under EU data protection law. In the final section, some guidance is provided to computer scientists to help them re-identify data in a way compliant with the GDPR. Some contentious points will be underlined.

2 Technical aspects related to re-identification

In order to understand why re-identification is paramount to improve security measures implemented on personal data, a quick background on re-identification is needed.

---

[17] Ibid.

[18] Esayas S. Y., The role of anonymisation and pseudonymisation under the EU data privacy rules: beyond the 'all or nothing' approach., European Journal of Law and Technology, Vol 6, No 2, 19, (2015).

[19] Finck, M., Pallas, F.: They who must not be identified—distinguishing personal from non-personal data under the GDPR., International Data Privacy Law, 2020, Vol. 10, No. 1, 35 (2020).

[20] Stalla-Bourdillon, S., Knight, A.: Anonymous Data V. Personal Data—A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data., International Data Privacy Law, Vol. 10, No. 1., (2020).

Over the years, re-identification attacks have become increasingly sophisticated and effective. In the current landscape, any de-identified data can potentially be subject to re-identification.[21] This is primarily due to the widespread availability and accessibility of data sources and online datasets that contain vast amounts of personal data. As previously mentioned, it is important to note that re-identification attacks have capitalised on progress in machine learning and other AI applications,[22] so that their potential and effectiveness cannot be overstated. Data controllers must be prepared to address these threats and the resulting consequences, including data breaches and risks to individuals' privacy.

Re-identification attacks have been successful on various kinds of data, such as health data, movie preferences, location data, university courses and users' search queries (see Table 1). Other techniques have been deployed to enable the free use of data without privacy risks, including synthetic data, but even in the latter case, data can be re-identified.[23] The nature of data is therefore irrelevant to study re-identification since no de-identified data is immune from re-identification attacks.

The increasing ease of cross-referencing data has changed how re-identification attacks are carried out. Cross-referencing refers to the process of comparing information from one dataset with information from another dataset: this information within these datasets is compared to identify matches.[24] This comparison can be based on specific identifiers or attributes. Datasets containing personal information often include various elements alongside the actual data, such as direct identifiers (e.g., social security numbers), indirect identifiers (e.g.; ZIP codes), and quasi-identifiers (attributes such as ZIP code, birthdate and gender).[25] While a direct identifier permits to uniquely identify an individual without additional knowledge, an indirect identifier permits such identification when combined with additional information. A quasi-identifier cannot uniquely identify an individual, but it is sufficiently well correlated with the individual. A set of quasi-identifiers can constitute a profile which uniquely identifies the individual. By cross-referencing a de-identified dataset with other sources, an adversary can obtain a set of quasi-identifiers that uniquely identifies an individual, potentially revealing her/his true identity/name.

The increasing availability of data sources for cross-referencing purposes has enabled anyone to use online and freely available data (for example, from social networks[26] such as IMDB or LinkedIn) to re-identify individuals in large datasets. Data openness and cross-referencing techniques both emphasise the following statement of the NIST: "[...]it is not possible to algorithmically determine what kinds of contextual information can be used to assist in future re-identification efforts.[27]" In other words, it is difficult to predict the capabilities of an adversary to undertake a re-identification attack in the future.

---

[21] Ibid.

[22] Rocher, L. et al.: Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models., Nature communications, (2019).

[23] Giomi, M. et al., A Unified Framework for Quantifying Privacy Risk in Synthetic Data., (2022).

[24] Yang, H., Yi, D., Liao, S., Lei, Z., & Li, S., Cross Dataset Person Re-identification. In ACCV Workshop., (2015).

[25] Garfinkel, S., De-Identification of Personal Information, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD, (2015).

[26] 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, 1376, (2013).

[27] Garfinkel, S., De-Identification of Personal Information, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD, (2015).

When a re-identification attack occurs, three types of information can be disclosed:[28] identity, attribute and inferred information. Identity disclosure occurs when an attacker successfully links de-identified data to a specific individual, directly revealing their identity. Attribute disclosure occurs when the adversary can attribute a piece of information to an individual without necessarily knowing their identity, thereby revealing personal attributes associated with that individual. Inferential disclosure occurs, according to the NIST, "when information can be inferred with high confidence from statistical properties of the released data[29]" providing insights into sensitive details about individuals even without direct identification. It is crucial to note that re-identification encompasses more than just identity disclosure, and all forms of disclosure, including attribute and inferential, must be carefully considered when assessing the risks associated with re-identification attacks, as they can lead to the processing of personal data.

Re-identification techniques grow more sophisticated alongside advancements in algorithms and availability of diverse data sources for cross-referencing. This risk extends to various forms of data, such as statistics, aggregated data,[30] and even machine learning models,[31] where individuals might be linked back to the initial data.

Furthermore, re-identification attacks also benefit from advancements in artificial intelligence. Models can be specifically trained to re-identify individuals, by singling them out or by performing inference attacks, exploiting datasets. Attacks based on machine learning have been carried out on medical data[32] or connection data.[33]

We can also speculate that quantum computers will also enhance the possibilities and effectiveness of re-identification attacks due to their ability to break traditional encryption methods through advanced algorithms.[34] The main technology used to secure data is cryptography (and particularly encryption). Cryptography is a fundamental aspect of privacy-enhancing technologies (PET), and the evolution of quantum computers may further augment the capabilities and efficacy of re-identification attacks, potentially presenting new challenges to data privacy and security.

The following table aims to give some insights into the most emblematic re-identification attacks.

---

[28] Ibid.

[29] Ibid.

[30] Willemson, J. (2022). Fifty Shades of Personal Data – Partial Re-identification and GDPR. In: Gryszczyńska, A., Polański, P., Gruschka, N., Rannenberg, K., Adamczyk, M. (eds) Privacy Technologies and Policy. APF 2022. Lecture Notes in Computer Science, vol 13279. Springer, Cham (2022).

[31] Shokri, R.: Membership Inference Attacks Against Machine Learning Models. (2017).

[32] Rocher, L. et al.: Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models, Nature communications, 10, 3069 (2019).

[33] De Montjoye, Y.-A., et al,.: Unique in the Crowd: The Privacy Bounds of Human Mobility. 3, 1376 (2013).

[34] European Union Agency for Network and Information Security, https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design, last accessed 2024/06/12.

Table 1. The rise of re-identification attacks.

Year	Dataset creator	Type of data	Attack type	References
1997	NAHDO GIC Cambridge Massachusetts	Hospitalisation records Medical records in the GIC data Voter registration data	Crossing databases	[35]
2000	NAHDO Cambridge Massachusetts	Hospitalisation records Voter registration data	Crossing databases (census)	[36]
2006	AOL	Users' search queries	Crossing databases	[37]
2007	Netflix	Users' movie preferences	Crossing databases	[38]
2008	Cabspotting	Taxi trajectory (GPS coordinates)	Point of interests discovery	[39]

[35] Sweeney, L., K-ANONYMITY: A MODEL FOR PROTECTING PRIVACY., International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 557-570 (2002)

[36] Sweeney, L., Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh (2000).

[37] TheNewYorkTimes, https://www.nytimes.com/2006/08/09/technology/09aol.html, last accessed 2023/06/09.

[38] Narayanan, A., Shmatikov, V., How To Break Anonymity of the Netflix Prize Dataset. (2006).

2017-2018	Strava	Users' trajectories (GPS coordinates)	Regression	[40]
2021	edX (Harvard)	Students enrolled in edX courses	Crossing databases	[41]

The table shows how re-identification attacks are exploiting additional data obtained by different means. Early examples include Latanya Sweeney re-identifying a governor using voter records and hospital data.[42] More recently, companies like AOL[43] and Netflix[44] released search history and movies ratings, that, when combined with other information, enabled people to be re-identified. Furthermore, the EdX incident exemplifies the evolving threat of re-identification attacks. Even with supposedly advanced de-identification methods, researchers were still able to re-identify users. These cases highlight the fact that using de-identification techniques and removing identifying information in order to protect data might not be enough to ensure data protection.[45]

Furthermore, most of the re-identification attacks cited in Table 1 have been performed by academics. Public research has a leading role in making progress in re-identification.

Additionally, in the realm of data security, parallels can be drawn between re-identification/anonymisation and cryptanalysis/cryptography. Just as cryptanalysis challenges cryptographic methods to enhance security protocols, re-identification efforts test anonymisation techniques to improve data privacy. Both domains benefit from this ongoing tension: vulnerabilities exposed through re-identification or cryptanalysis lead to stronger anonymisation and cryptographic methods. For instance, early cryptographic methods relied heavily on simple ciphers, which only shifted letters by a fixed number of positions. While these early methods provided basic encryption, they were relatively easy to break with simple frequency analysis.

---

[39] Gambs, S et al., De-Anonymization Attack on Geolocated Data., Journal of Computer and System Sciences, 80, 1597 (2014).

[40] Dhondt K., et al., A Run a Day Won't Keep the Hacker Away: Inference Attacks on Endpoint Privacy Zones in Fitness Tracking Social Networks, CCS '22: Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, 801–814 (2021).

[41] Cohen, A., Attacks on Deidentification's Defenses. 31st, USENIX Security Symposium (USENIX Security 22), 1469–1486, (2022).

[42] Sweeney, L.: K-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. ., International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10 (5), 557-570 (2002).

[43] Tech Crunch, https://techcrunch.com/2006/08/06/aol-proudly-releases-massive-amounts-of-user-search-data/?guccounter=1, last accessed 2024/04/29

[44] Wired, https://www.wired.com/images_blogs/threatlevel/2009/12/doe-v-netflix.pdf, last accessed 2023/01/25.

[45] Mitchum, R., New Kind of Attack Called "Downcoding" Demonstrates Flaws in Anonymizing Data. (2022).

Over time, as cryptanalysts uncovered these weaknesses, more advanced encryption techniques were developed, which use complex mathematical algorithms. It is important to note that while encryption is not an anonymisation technique, it can serve as a powerful pseudonymisation tool.[46] Re-identification thus remains a potential risk if the encryption key is compromised.

As for re-identification, early anonymisation techniques were only removing direct identifiers in datasets (see the AOL and Netflix examples). Re-identification has proven that it was not enough to prevent an adversary from re-identifying the individuals from those weakly anonymised datasets. Just as cryptanalysis has demonstrated that basic cryptographic methods are insufficient for protecting sensitive data, re-identification has shown the inadequacy of early anonymisation techniques. Both fields thrive on this continuous push and pull, as the advancements in one drive improvements in the other. Nonetheless, more efforts were needed to reach a stronger form of anonymisation. Re-identification is thus still needed today to evaluate new anonymisation proposals.[47]

State-of-the-art re-identification techniques present a tricky dilemma for anonymisation efforts. Enhancing anonymisation also means enhancing re-identification, which may seem paradoxical and counter-intuitive. Without effective anonymisation, individuals' privacy is compromised. However, robust re-identification techniques are also necessary to prevent re-identification attacks on anonymised data. Understanding the legal framework becomes essential to navigate this delicate balance and to preserve the privacy of personal data, within the confines of existing regulations.

3 What EU law says about re-identification

As a first step, it is worth emphasising what data protection law says about re-identification.

3.1 Prohibition of re-identification under the Data governance act

While the GDPR does not explicitly prohibit re-identification, it emphasises data protection principles that implicitly discourage practices leading to re-identification. However, other legal frameworks in the EU do prohibit re-identification.

For instance, the newly adopted Data governance act provides that "[r]e-identification of data subjects from anonymised datasets should be prohibited,[48]" which suggests that the situation regarding the GDPR's treatment of re-identification could change in the future, but, at the moment, it is not explicitly addressed within the GDPR itself.

This same regulation also provides that "[r]e-users shall be prohibited from re-identifying any data subject to whom the data relates and shall take technical and operational measures to pre-

[46] European Data Protection Supervisor, https://www.edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf, last accessed 2024/06/13.

[47] Kikuchi H., et al. Ice and Fire: Quantifying the Risk of Re-identification and Utility in Data Anonymization, 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), Crans-Montana, Switzerland, 1035-1042 (2016).

[48] See recital 8 of the REGULATION (EU) 2022/868 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act).

vent re-identification and to notify any data breach resulting in the re-identification of the data subjects concerned to the public sector body.[49]" In the latter provision, the prohibition of re-identification seems to be more strictly defined. However, such a conclusion must not be overestimated, as it is mainly designed to frame the situation where a public body shares data with another party, following certain procedures. In this scenario, it is obvious that one of the conditions for sharing data is that the recipient of data will not try to re-identify them. This provision does not address a classic re-identification scheme, where poorly anonymised data are published online and later re-identified by an organisation, or a natural person.

It is worth mentioning that the DGA covers both personal and non-personal data, with the GDPR applying whenever personal data is involved.[50] The regulation emphasises the need for data interoperability and protection against re-identification while encouraging the development of secure data processing environments and standardised anonymisation techniques, which will likely enhance the ability to share data safely and reduce the risk of re-identification. It prohibits re-identification of data subjects from anonymised datasets. However, it is essential to notice that the DGA is not a data protection regulation but rather a legal framework designed to promote data sharing and reuse within the EU.

3.2 Identifiability under EU Data protection law

Identifiability and re-identification, though conceptually distinct, often yield similar practical outcomes concerning privacy risks. Both identifiability and re-identification pose significant privacy risks, as they can lead to the exposure of personal information. On the one hand, under the GDPR, if data can be linked to an individual, it is considered personal data. The focus is on whether a person can be identified, directly or indirectly, from the data in question. On the other hand, re-identification involves transforming de-identified data back into identifiable data. Thus, while identifiability primarily addresses the inherent link between data and individuals, re-identification underscored the vulnerability of supposedly protected datasets, by demonstrating the potential for originally de-identified data to be transformed into identifiable information. The re-identification process directly influences the identifiability of data.

The European Court of Justice (ECJ) gives us some clues into the issue of re-identification and identifiability under EU data protection law. Indeed, in *Breyer*, the Court assesses whether an IP address can be regarded as personal data for a web service provider. The Court therefore had to evaluate whether the said web service provider had "reasonable means" to identify data, as provided for by Recital 26. While the terminology may differ, the evaluation of identifiability addresses similar concerns as re-identification, as both have the same effects.

From this background the Court claims that the means would not be reasonably likely to be used if "the identification of the data subject was prohibited by law.[51]" This criterion has been recalled by the General Court in case T 557/20 which involved two EU organs and institutions,

---

[49] See Article 5 (5) of the Data Governance Act.

[50] European Commission, https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained, last accessed 2024/06/12.

[51] ECJ, JUDGMENT OF THE COURT (Second Chamber), in case C-582/14, Patrick Breyer v. Bundesrepublik Deutschland, 19 October 2016, § 46.

namely the Single Resolution Board and the EDPS.[52] Essentially, this suggests that if re-identification is made unlawful, said re-identification cannot be deemed reasonable and that the assessment of the means reasonably likely to be used becomes irrelevant. The prohibition of re-identification is thus considered as a silver bullet, allegedly representing the most effective means to protect personal data. It signifies that re-identification cannot be considered reasonable when it contravenes legal restrictions. Due to the absence of a general prohibition of re-identification under the GDPR, the question shifts to whether such prohibitions exist at a national level.

3.3 Approaches to re-identification

Some states have implemented specific regulations and guidelines to address re-identification within their jurisdictions. For instance, the UK has expressly incorporated a provision in the Data Protection Act that prohibits re-identification attacks. Indeed, in a chapter dedicated to "offences relating to personal data", section 171 provides that "[i]t is an offence for a person knowingly or recklessly to re-identify information that is de-identified personal data without the consent of the controller responsible for de-identifying the personal data.[53]" Interestingly, UK law prohibits re-identification as a process, without putting the emphasis on the means by which such re-identification occurs. This provision sets some exceptions to this general rule such as re-identification carried out for public interest reasons, in particular for research purposes. In the explanatory notes of the bill, UK lawmakers underline that this provision tackles the issue of de-identified data published online, in particular when they are health data which can lead to the re-identification of patients.[54] It must be emphasised here that there have been huge controversies in the UK with regard the re-identification of doctors who have carried out late-term abortion from statistics released by the department of health.[55] Judicial authorities upheld that statistics were not personal data.[56]

Although the GDPR no longer applies in the UK, the principles and legal interpretations remain relevant for understanding how member states might handle re-identification prohibition and how they might evaluate the lawfulness of re-identification means and methods.

3.4 Evaluating the possibility of criminalising re-identification

With regard to EU law, recital 149 of the GDPR enables Member states to "lay down the rules on criminal penalties for infringements of this Regulation, including for infringements of na-

[52] CJEU, JUDGMENT OF THE GENERAL COURT (Eighth Chamber, Extended Composition) in Case T-557/20, Single Resolution Board (SRB) v. European Data Protection Supervisor (EDPS), 26 April 2023.

[53] UK Public General Acts, 2018 c.12, Data Protection Act 2018, legislation.gov.uk

[54] Data Protection Act 2018, Explanatory Notes, Commentary on provisions of the act, § 492.

[55] The Guardian, https://www.theguardian.com/society/2009/oct/16/pro-life-alliance-abortion-jepson-case, last accessed 2024/04/29.

[56] Department of Health, R (on the application of) v. Information Commissioner, England and Wales High Court (Administrative Court), 20 April 2011.

tional rules adopted pursuant to and within the limits of this Regulation.[57]" One should thus conclude from this provision that it is up to member states to criminalise (or not) re-identification by adopting domestic laws addressing this issue. The EU thus leaves to EU member states the final decision as to render re-identification unlawful.[58] However, this does not imply that the CJEU offers no guidance into the potential unlawfulness of re-identification schemes.

3.5 Evaluating the unlawfulness of re-identification means

In *Breyer*, the Court underlined – as it has been previously stated - that lawfulness is a criterion to be taken into account when assessing the reasonable means likely to be used to re-identify data. The Advocate General even noted that "[i]t is irrelevant, in that context, that access to the personal data is possible *de facto* by infringing data protection laws.[59]" What is interesting here is that the Advocate General seems to consider that being able in practice to re-identify data may be a violation of data protection law: within the context of Directive 95/46 (the EU Data Protection Directive), the practical possibility of accessing personal data must be considered reasonable only if it is done through lawful means. In other words, any means of access to personal data must comply with applicable data protection laws and regulations.

The Advocate General emphasises that the requirement for access to be reasonable inherently implies that it must be lawful. This means that even if there are practical methods to access personal data, such access would not be considered reasonable if it involves infringing data protection laws. It is thus irrelevant whether access to personal data is possible in practice through methods that violate data protection laws. Even if such unauthorised access methods exist, they cannot be considered a reasonable means of access under Directive 95/46.

Second, the Court emphasises that the legal means are manifested by the existence of legal channels, which supposes that there must be legal provisions allowing a specific person to get the information needed to re-identify data.

Eventually, the Court is interested in the means used to re-identify, and not by the re-identification by itself. This solution has been reiterated by the General Court[60] in the SRB vs EDPS case.[61]

From this perspective there are still uncertainties with regard to this "legal means" criterion: does it mean that there are unlawful means? If so, what are they? Does it mean that, positively,

---

[57] See recital 149 of the GDPR.

[58] It is important to acknowledge that certain practices can be deemed unlawful without necessarily falling under the scope of criminal law. This acknowledgment is particularly relevant in the context of the GDPR, which primarily emphasises administrative and civil measures to ensure data protection. Under the GDPR, the focus extends beyond criminal prohibitions to encompass a broader legal framework. For example, a breach of data protection principles, such as inadequately anonymising personal data, can result in significant administrative fines and sanctions imposed by data protection authorities.

[59] CJEU, OPINION OF ADVOCATE GENERAL CAMPOS SÁNCHEZ-BORDONA delivered on 12 May 2016, Case C-582/14 Patrick Breyer v. Bundesrepublik Deutschland.

[60] CJEU, JUDGMENT OF THE GENERAL COURT (Eighth Chamber, Extended Composition), in Case T-557/20, Single Resolution Board (SRB) v. European Data Protection Supervi- sor (EDPS), 26 April 2023.

[61] Lodie A., Are personal data always personal? Case T-557/20 SRB v. EDPS or when the qualification of data depends on who holds them., European Law Blog, (2023).

there must be legal channels, explicit provisions that enable us to collect additional information? In other words, is re-identification expressly allowed or at least possible?

One might even wonder whether, in practice, the "lawfulness" of re-identification is really a relevant criterion to increase the level of protection of personal data. From this perspective, the Swedish Data Protection Authority (DPA) has cautioned against interpreting the concept of personal data in a manner that excessively limits the scope of protection, as it would significantly weaken the overall protection offered by the GDPR. The Swedish DPAs view can be read as follows:

"[a]n interpretation of the concept of personal data that means that it must always be demonstrated that there is a legal possibility to link such data to a natural person would, according to IMY, entail a significant limitation of the regulation's scope of protection, and open up opportunities to circumvent the protection in the regulation. This interpretation would, among other things, be contrary to the purpose of the regulation as set out in Article 1(2) of the GDPR.[62]"

Specifically, the Swedish DPA's interpretation suggests that data can be considered anonymised—and thus not subject to GDPR restrictions—even if there is a risk that the data can be re-identified using sophisticated techniques. This broad interpretation allows for greater access to such "anonymised" data, which could then be used in ways that might not fully protect individuals' privacy.

The main concern here is about the robustness of the anonymisation standards being applied. If data that can be re-identified is still treated as anonymised, the protections that the GDPR aims to provide might be undermined. This means that individuals' personal information could be exposed or misused, despite the intention to keep it private. The core issue is that re-identification techniques are advancing, and what might be considered anonymised today could become identifiable tomorrow. Therefore, the interpretation by the Swedish DPA raises questions about whether current anonymisation practices are sufficient to safeguard personal data in the long term.

This would impose a strict requirement that may exclude certain types of data from being considered personal data, even if they pose potential risks to individuals' privacy. Entities may exploit loopholes by structuring their data practices in a way that avoids meeting the strict legal criteria for personal data, as they would all outside the scope of the GDPR.

Consequently, the purpose of the following section will be precisely devoted to analysing the compliance of re-identification schemes with the GDPR. More specifically, as underlined previously, we will try to figure out whether re-identification for research purposes can be deemed lawful under EU data protection law just like UK lawmakers expressly enshrined.

4 Re-identification under the GDPR in practice

In this section we will try to provide some insights on the way re-identification schemes can be deemed compliant with the GDPR. Although it does not constitute a handbook for practitioners or researchers willing to carry out re-identification, we underline some contentious points that should be taken into account.

[62] IMY, Supervisory decision under the General Data Protection Regulation Tele2 Sverige AB's transfer of personal data to third countries, DI-2020-11373, 30 June 2023.

4.1 Re-identification as data processing

The first question that one has to address is whether re-identification is subject to the GDPR, or more generally, to EU data protection law. The GDPR applies materially to "the processing of personal data.[63]" Re-identification, to be subject to the GDPR, must constitute data processing.

Under the GDPR, data processing involves (among other actions) collecting, consulting or using data.[64] From this perspective, carrying out re-identification research should be considered as data processing since researchers will at least consult and store the data once re-identified. The purpose of this data processing operation lies in the scientific progress that computer scientists accomplish by discovering new weaknesses of a de-identification technique used to release a dataset publicly or to protect data.

From this background, the Norwegian DPA claims for instance that "(i)f someone should succeed in re-identifying the data, and this results in personal data being processed, the organisation responsible for the data must assume the role of data controller for them.[65]" Re-identification, and the subsequent storing, sharing or re-use of data must be considered data processing for which the organisation re-identifying data assumes the role of data controller.

This conclusion seems to be in line with Article 4 (7) of the GDPR[66] since researchers re-identifying data determine the purposes and means of data processing. Indeed, they use re-identification tools to reveal anonymisation vulnerabilities which may lead to massive data breaches. The organisation the researcher works for could be considered as a data controller in this context as well, as the French DPA (CNIL[67]) suggested.[68]

For instance, the University employing a computer scientist to work on re-identification issues can be considered as a data controller, but we will not discuss this issue further.

When a researcher re-identifies pseudonymised data, it is obvious that data are being processed since said researcher receives personal data (pseudonymised data) and further processes them in order to re-identify them.

4.2 Contentious points relating to re-identification for research purposes with regard to data processing principles

Since computer scientists re-identifying data should logically be considered as data controllers,[69] they must comply with at least one of the legal bases provided for in the GDPR. Indeed, the first principle relating to data processing as laid down in Article 5 of the GDPR is that "data

---

63 See Article 2 of the GDPR.

64 See Article 4 (2) of the GDPR.

65 Datatilsynet, https://www.datatilsynet.no/en/regulations-and-tools/reports-on-specific-subjects/anonymisation/?print=true, last accessed 2023/01/25.

66 See Article 4 (7) of the GDPR.

67 The French Data Protection Authority (CNIL) (Commission Nationale de l'Informatique et des Libertés) has been a critical player in the landscape of data protection, particularly with the enforcement of GDPR in France.

68 CNIL, https://www.cnil.fr/fr/recherche-scientifique-hors-sante-les-questions-reponses-de-la-cnil, last accessed 2023/01/25.

69 Cf above, subsection 4.1.

shall be (…) processed lawfully.[70]" The existence of a legal basis is not enough to process data lawfully pursuant to the GDPR but it remains a salient issue since it reveals all the difficulties that may arise when it comes to re-identification.

As a preamble, it is worth mentioning that in this situation, the purpose of the data processing operation is a scientific purpose since the main aim is to reveal data security vulnerabilities and thus protect data subjects' personal data and privacy. However, the scientific purpose is not, by itself, a legal basis. In other words, processing data for research purposes, or any other "legitimate" purpose does not mean that such processing is lawful under the GDPR or benefits from a legal basis.

From this background, the French DPA released guidelines on the legal regime applicable to data processed for scientific purposes. The guidelines identify as possible legal bases, the consent of subjects, the performance of a task carried out in the public interest, or the legitimate interest pursued by the controller.[71] However, we will see that each of these candidates may involve interpretation issues or do not fit the reality of re-identification attacks carried out for scientific purposes.

First, consent is not likely to be a relevant legal basis for processing data when launching a re-identification attack. As a matter of fact, researchers carrying out such an attack are unaware of who the data subjects were initially, since the main aim of their operation is to try to identify data subjects from de-identified datasets. Such a legal basis is thus inoperative to provide a clear framework.

CNIL's guidelines also mention the performance of a task carried out in the public interest. However, once again, it is unclear whether such a legal basis is fit for purpose in such a scenario. More specifically, the GDPR requires that "(w)here processing is carried out in accordance with a legal obligation to which the controller is subject or where processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority, the processing should have a basis in Union or Member State law.[72]" In other words, member states' law should expressly contain provisions regarding data processing carried out by researchers willing to re-identify data. The question goes as to whether this law must be specific or whether a general statute of researchers under domestic law could be sufficient as well as a mere transposition of the GDPR into domestic law.[73]

For instance, the French "Loi informatique et libertés" transposing the GDPR into French law contains some provisions which can be interesting when considering the legal basis to process data in the public interest. Article 78 of the law provides that "A decree [...] shall determine under what conditions and subject to what safeguards the rights provided for in Articles 15, 16, 18 and 21 of the same Regulation may be waived in whole or in part with regard to processing for scientific or historical research purposes, or for statistical purposes.[74]" Said decree has been adopted and it interestingly provides that when processing data for scientific

[70] See article 5 of the GDPR.

[71] CNIL,https://www.cnil.fr/sites/cnil/files/atoms/files/consultation_publique_presentation_du_regime_juridique_applicable_aux_traitements_a_des_fins_de_recherche.pdf, last accessed 2023/01/25.

[72] See recital 45 of the GDPR.

[73] See for instance French Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, art 78 and Décret n° 2018-687 du 1er août 2018 pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, modifiée par la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles, art. 100-1.

[74] Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

purposes, data controllers and processors shall "respect the rules of ethics applicable to their sectors of activity.[75]" Although such texts do not expressly state that computer scientists carrying out re-identification attacks for scientific purposes can rely upon the public interest legal basis, they are worth mentioning since they deal with the obligations of data controllers in the field of research.

It is undeniable that computer scientists performing a re-identification attack contribute to the progress of research in the field of cybersecurity and data protection. As such, researchers, as members of a public-funded institution, placed under the authority of the Ministry of Higher Education and Research, can be considered as exercising a task carried out in the public interest.

The last legal basis likely to authorise researchers to undertake a re-identification attack is the legitimate interest of the data controller. However, this legal basis seems to be inoperative for our case study since "this basis applies only to private entities.[76]" Indeed, recital 47 of the GDPR provides that "(g)iven that it is for the legislator to provide by law for the legal basis for public authorities to process personal data, that legal basis should not apply to the processing by public authorities in the performance of their tasks.[77]"

Actually, things are much more nuanced since it depends on the legal status of universities in EU Member States. Can a university be considered a public authority? As already mentioned, this is the case in France, but the status of researchers and research bodies may vary from one EU Member State to another. Besides, even when researchers are employed by a public authority, legitimate interest would be excluded as a valid legal basis only when processing is carried out in the performance of their tasks. One may consider that, since research is the core function of researchers and research public bodies such as universities, the legitimate interest would not be a valid legal basis in the context of computer scientists undertaking re-identification attacks for research purposes.

The French DPA has clarified the use of the "legitimate interest" legal basis by public bodies to process data. It underlined that "[t]he GDPR provides that the legal basis of legitimate interest does not apply to processing carried out by public authorities in the performance of their tasks". "However, [...] this provision does not prevent the use of this legal basis when the processing, although necessary for its current administration or operation, does not fall within the strict performance of its tasks as provided for by the texts.[78]" This tends to exclude legitimate interest from the scope of the legal bases likely to be used for research in the field of re-identification.

Interestingly, some universities have published their own guidelines to specify under what grounds their agents could process data. For instance, the University College London (UCL) claims on its official website that "(a)s a public authority, most of UCL's processing will be undertaken using Article 6(1)(e) above, the 'public task' condition. This applies when the processing is necessary for UCL to perform a task in the public interest. Examples include most of UCL's research, teaching and learning activities – we can clearly demonstrate a 'public task' basis for these because performing such tasks is a core part of UCL's Charter and Statutes".[79]

---

[75] Décret n° 2019-536 du 29 mai 2019 pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

[76] Maldoff G., How GDPR changes the rules for research, (2016).

[77] See Recital 47 of the GDPR.

[78] CNIL, https://www.cnil.fr/fr/les-bases-legales/choisir-base-legale, last accessed 2023/04/13.

[79] University College London, Practical Data Protection Guidance Notice: Legitimate interests as a lawful basis for processing personal data, https://www.ucl.ac.uk/data-

In the same line of thought, the French "CNRS" (National Centre for Scientific Research) published guidelines on GDPR compliance in the field of research. These guidelines are quite similar to the UCL ones since the CNRS claims that "[i]n the context of research activities, processing should preferably be carried out on the basis of consent (respecting the principle of informational self-determination), but processing may also be based on the exercise of a public interest mission.[80]"

The 'public interest' legal basis seems, therefore, to be relevant when considering the re-identification of anonymised datasets carried out by computer scientists in the fulfilment of their tasks.

However, the lawfulness of processing is not the only core principle which is likely to raise issue when it comes to re-identification.

4.3 Re-identification with regard to GDPR's core principles: Transparency and Data minimisation

One may also question the compliance of re-identification attacks as regards the principle of transparency as provided for in Article 5 of the GDPR.[81] Re-identification is by its very nature a covert data processing operation since, when processing data, researchers do not know who the data belong to, so they cannot be transparent on the way they process data vis-à-vis data subjects.

In addition to the transparency principle, re-identification challenges the principle of data minimisation as well. This principle mandates that data controllers limit the collection, processing, and retention of personal data to what is strictly necessary for the intended purpose. However, this principle may conflict with the objectives of re-identification attacks conducted for scientific research, which aim to uncover vulnerabilities in anonymisation techniques by analysing extensive datasets. As a matter of fact, the very idea of re-identification involves cross-referencing data by collecting vast amounts of data to single out data subjects, to link some attributes to them or infer information.[82]

4.4 Re-identification and data subjects' rights

One of the main issues with regard to the compliance of re-identification attacks with the GDPR is the exercise of data subjects' rights. Data subjects have a right to be informed of the processing of their data, a right to object to such processing or even in some situations a right to erasure.

In particular, in the scenario that we are considering, it seems very difficult for data controllers (researchers) to comply with the right of data subjects to information as provided for by

protection/guidance-staff-students-and-researchers/practical-data-protection-guidance-notices/legitimate, last accessed 2023/01/25.

80 InSHS IAP and others, Les Sciences Humaines et Sociales et La Protection des Données à Caractère Personnel Dans Le Contexte de La Science Ouverte: Guide Pour La Recherche. (2023).

81 See Article 5 of the GDPR.

82 Yang, H., Yi, D., Liao, S., Lei, Z., & Li, S., Cross Dataset Person Re-identification. In ACCV Workshop., (2015).

Article 14 of the GDPR.[83] Indeed, when re-identifying an anonymised dataset, researchers do not know who the data subjects are, so they cannot inform them about the data processing carried out.[84] They can only inform them *a posteriori*, which is not the right way to proceed since "(n)otice should be provided at the time when the data is first collected, and it must include the controller's identity and contact information.[85]"

However, Article 14 paragraph 5 provides for some exemptions, in particular, data controllers do not have to provide a notice when the situation makes it impossible or too complex. The same goes when such a requirement is likely "to render the processing impossible or seriously impair the achievement of the objectives of that processing.[86]"

By virtue of this article, computer scientists undertaking re-identification attacks would not be constrained to inform people since it would be impossible as they do not know the exact nature of the data processed, nor who the data subjects actually are.

Furthermore, under the GDPR, data subjects benefit from other rights concerning the processing of their data.[87] However, when data processing for scientific purposes is involved, data controllers may be exempted from compliance with these obligations.[88] It means that researchers undertaking a re-identification attack would not have to protect all these rights. Nonetheless, these exemptions must be provided by European or domestic law, besides they are not absolute and "must be necessary for the fulfilment of [the research] purposes.[89]" Indeed, these derogations must be interpreted narrowly and the research project must comply with the GDPR in other respects.[90]

---

[83] Article 14 of the GDPR reads as follows:

'1. Where personal data have not been obtained from the data subject, the controller shall provide the data subject with the following information:

(a) the identity and the contact details of the controller and, where applicable, of the controller's representative;

(b) the contact details of the data protection officer, where applicable;

(c) the purposes of the processing for which the personal data are intended as well as the legal basis for the processing;

(d) the categories of personal data concerned;

(e) the recipients or categories of recipients of the personal data, if any;

(f) where applicable, that the controller intends to transfer personal data to a recipient in a third country or international organisation and the existence or absence of an adequacy decision by the Commission, or in the case of transfers referred to in Article 46 or 47, or the second subparagraph of Article 49(1), reference to the appropriate or suitable safeguards and the means to obtain a copy of them or where they have been made available.'

[84] For instance, computer researchers cannot reasonably inform data subjects in a research context when they analyse historical medical records studying effects of certain treatments from last century. While data subjects may still be alive today, locating and contacting them individually is practically impossible due to the incomplete nature of the records, changes in contact information etc.

[85] Maldoff G., How GDPR changes the rules for research, (2016).

[86] See Article 14 § 5 (b) of the GDPR.

[87] See in particular Article 15, 16, 17, 18, 20, 21 of the GDPR.

[88] See Article 89 of the GDPR.

[89] Maldoff G., How GDPR changes the rules for research, (2016).

[90] Office of the Data Protection Ombudsman, Rights of the data subject in scientific research, https://tietosuoja.fi/en/rights-of-the-data-subject-in-scientific-research

These rights include the right to access their personal data[91] held by data controllers, allowing them to verify the lawfulness of the processing. Additionally, data subjects have the right to rectify inaccurate or incomplete personal data,[92] ensuring the information held about them is accurate and up-to-date. Furthermore, individuals can request the erasure of their personal data under certain circumstances, commonly referred to as the "right to be forgotten."[93] They also have the right to restrict processing,[94] to object to processing, and the right to not be subject to a decision based solely on automated processing.[95]

Eventually, the way re-identification operates can contradict the philosophy of the GDPR which is to ensure privacy by design and by default.

4.5 Re-identification and Privacy by design and default

Similarly, privacy by design and default, another key aspect of the GDPR highlighted in Article 25, mandates that data protection measures be integrated into the design and operation of systems, ensuring that privacy is maintained by default. While data protection by design involves integrating privacy concerns during the whole lifecycle of a product or service,[96] privacy by default "refers to the selection of the most privacy friendly configuration by default."[97]. It is clear from what has been stated above that re-identification runs contrary to the very nature of the principles of privacy by design and privacy by default.

A re-identification scheme is by default aimed at cross-referencing and collecting as much information as possible to succeed in re-identifying data subjects. Such a system is invasive and intrusive by design and default.

These issues underscore the complexity surrounding the intersection of data protection principles and re-identification activities for scientific purposes. While re-identification may serve legitimate research objectives and public interest missions, reconciling these activities with fundamental data protection principles remains a legal grey area. Clarity is needed regarding the extent to which derogations from data subjects' rights are permissible for researchers conducting re-identification attacks and whether specific legal provisions or broader research statutes suffice to authorise such activities. Addressing these concerns is essential for ensuring compliance with the GDPR while facilitating valuable scientific research.

5 Conclusion

Re-identification has considerably developed over the years, posing significant risks for individuals' privacy and data protection. The proliferation of data-driven technologies and the widespread collection of personal information for public re-use and marketing purposes have

[91] See Article 15 of the GDPR.
[92] See Article 16 of the GDPR.
[93] See Article 17 of the GDPR.
[94] See Article 18 of the GDPR.
[95] See Article 22 of the GDPR.
[96] Jasmontaite L., Kamara I., Zanfir-Fortuna G., Leucci S., Data Protection by Design and by Default: Framing Guiding Principles into Legal Obligations in the GDPR. European Data Protection Law Review, 4, 2, 168–189 (2018).
[97] Ibidem.

intensified these privacy risks. While re-identification techniques offer insights into the effectiveness of data de-identification methods and risk mitigation strategies, they also raise complex legal questions, be it for data controllers, computer scientists, researchers and users. In particular, some of the requirements laid down by the GDPR seem to be hard to meet for researchers willing to carry out research in the field of re-identification.

To address these challenges, we propose several guidelines and future directions. Firstly, European Data Protection Authorities and institutions should release specific guidelines on the lawfulness of re-identification for research purposes, including defining the scope of permissible research activities and the conditions under which re-identification is lawful. Legally, clear definitions of research scope can help ensure that data usage is limited to what is necessary, minimising the risk of re-identification. Secondly, implementing robust technical measures to protect data, including encryption, anonymisation, pseudonymisation and access controls is crucial. Thirdly, developing and enforcing organisational policies could improve data protection and could ensure compliance with data minimisation and privacy by design and by default principles.

Acknowledgments. This work has been supported by the ANR 22-PECY-0002 IPOP (Interdisciplinary Project on Privacy) project of the Cybersecurity PEPR and by Inria action-exploratoire DATA4US. **The authors would like to warmly thank Cedric Lauradoux, researcher at Inria and general co-chair of the APF conference, for his insights on the technical part of the work and for his help.**

Disclosure of Interests. The authors are part of the Privatics team, Inria and work in close collaboration with Cédric Lauradoux, co-chair of the APF conference, on a regular basis.

References

1. Cohen, A., Attacks on Deidentification's Defenses. 31st, USENIX Security Symposium (USENIX Security 22), 1469–1486, (2022)
2. 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, 1376, (2013)
3. Dhondt K., et al., A Run a Day Won't Keep the Hacker Away: Inference Attacks on Endpoint Privacy Zones in Fitness Tracking Social Networks, CCS '22: Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, 801–814 (2021)
4. Esayas S. Y., The role of anonymisation and pseudonymisation under the EU data privacy rules: beyond the 'all or nothing' approach., European Journal of Law and Technology, Vol 6, No 2, 19, (2015)
5. Finck, M., Pallas, F.: They who must not be identified—distinguishing personal from non-personal data under the GDPR., International Data Privacy Law, 2020, Vol. 10, No. 1, 35 (2020)

6. Gambs, S et al., De-Anonymization Attack on Geolocated Data., Journal of Computer and System Sciences, 80, 1597 (2014)
7. Garfinkel, S., De-Identification of Personal Information, NIST Interagency/Internal Report (NISTIR), National Institute of Standards and Technology, Gaithersburg, MD, (2015)
8. Giomi, M. et al., A Unified Framework for Quantifying Privacy Risk in Synthetic Data., (2022)
9. Henriksen-Bulmer, J., Jeary, S.: Re-identification attacks—A systematic literature re-view, International Journal of Information Management, 36, 1184–1192 (2016)
10. Jasmontaite L., Kamara I., Zanfir-Fortuna G., Leucci S., Data Protection by Design and by Default: Framing Guiding Principles into Legal Obligations in the GDPR. European Data Protection Law Review, 4, 2, 168–189 (2018)
11. Kikuchi H., et al. Ice and Fire: Quantifying the Risk of Re-identification and Utility in Data Anonymization, 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), Crans-Montana, Switzerland, 1035-1042 (2016)
12. Lodie A., Are personal data always personal? Case T-557/20 SRB v. EDPS or when the qualification of data depends on who holds them., European Law Blog, (2023)
13. Lodie A., Case C-479/22 P, Case C-604/22 and the limitation of the relative approach of the definition of 'personal data' by the ECJ., European Law Blog, (2024)
14. Maldoff G., How GDPR changes the rules for research, (2016)
15. Mitchum, R., New Kind of Attack Called "Downcoding" Demonstrates Flaws in Anonymizing Data. (2022)
16. Mourby, M. et al.: Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK., Computer Law & Security Review, Vol 34, 222 – 233, (2018)
17. Narayanan, A., Shmatikov, V., How To Break Anonymity of the Netflix Prize Dataset. (2006).
18. Peloquin, D., DiMaio, M., Bierer, B. et al. Disruptive and avoidable: GDPR challenges to secondary research uses of data. Eur J Hum Genet 28, 697–705 (2020)
19. Rocher, L. et al.: Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models., Nature communications, (2019)
20. Rogers, J., Song A., Digital Marketing in The Legal Profession: What's Going On and Does It Matter?. Law, Technology and Humans, vol. 5, no. 2, 134 – 64 (2023)
21. Schermer, B.W., The Limits of Privacy in Automated Profiling and Data Mining. Computer Law & Security Review, (2011)
22. Sheth, J., Charles H., Next Frontiers of Research in Data Driven Marketing: Will Techniques Keep up with Data Tsunami?. Journal of Business Research, vol. 125, 780–84 (2021)
23. Shokri, R.: Membership Inference Attacks Against Machine Learning Models. (2017)
24. Spindler, G., Schmechel, P.: Personal Data and Encryption in the European General Data Protection Regulation., JIPITEC, 164 (2016)
25. Stalla-Bourdillon, S., Knight, A.: Anonymous Data V. Personal Data—A False Debate: An EU Perspective on Anonymization, Pseudonymization and Personal Data., International Data Privacy Law, Vol. 10, No. 1., (2020)
26. Sweeney, L., K-ANONYMITY: A MODEL FOR PROTECTING PRIVACY., International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 557-570 (2002)
27. Sweeney, L., Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh (2000)

28. Willemson, J. (2022). Fifty Shades of Personal Data – Partial Re-identification and GDPR. In: Gryszczyńska, A., Polański, P., Gruschka, N., Rannenberg, K., Adamczyk, M. (eds) Privacy Technologies and Policy. APF 2022. Lecture Notes in Computer Science, vol 13279. Springer, Cham (2022)
29. Yang, H., Yi, D., Liao, S., Lei, Z., & Li, S., Cross Dataset Person Re-identification. In ACCV Workshop., (2015)
30. Zuiderveen Borgesius, F., The Breyer Case of the Court of Justice of the European Union: IP Addresses and the Personal Data Definition., European Data Protection Law Review, Vol 3, Issue 1, 130–137 (2017)