Privacy-Enhancing Technologies and Anonymisation in Light of GDPR and Machine Learning

Fischer-Hübner, Simone; Hansen, Marit; Hoepman, Jaap Henk; Jensen, Meiko

Published in:
Privacy and Identity Management

DOI:
10.1007/978-3-031-31971-6_2

Privacy-Enhancing Technologies and Anonymisation in Light of GDPR and Machine Learning

Simone Fischer-Hübner[1], Marit Hansen[2], Jaap-Henk Hoepman[1,3,4], and Meiko Jensen[1(✉)]

[1] Karlstad University, Karlstad, Sweden
{simone.fischer-huebner,meiko.jensen}@kau.se
[2] Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, Kiel, Germany
marit.hansen@datenschutzzentrum.de
[3] Radboud University, Nijmegen, The Netherlands
jhh@cs.ru.nl
[4] University of Groningen, Groningen, The Netherlands
j.h.hoepman@rug.nl

Abstract. The use of Privacy-Enhancing Technologies in the field of data anonymisation and pseudonymisation raises a lot of questions with respect to legal compliance under GDPR and current international data protection legislation. Here, especially the use of innovative technologies based on machine learning may increase or decrease risks to data protection. A workshop held at the IFIP Summer School on Privacy and Identity Management showed the complexity of this field and the need for further interdisciplinary research on the basis of an improved joint understanding of legal and technical concepts.

1 Introduction

The European General Data Protection Regulation (GDPR) regulates the processing of personal data. Anonymised data does not fall under its legal regime (cf. Recital 26 of the GDPR, [1]). While the GDPR does not define the concept of "anonymisation", Recital 26 clarifies that "anonymous information"—e.g. as a result of the process of anonymisation—is information which does not relate to an identified or identifiable natural person. What does this mean? Recital 26 explains: "To determine whether a natural person is identifiable, 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."

The naive application of technical and organisational measures, specifically of so-called "anonymisation technologies", aiming at a successful anonymisation of

Published by Springer Nature Switzerland AG 2023
F. Bieker et al. (Eds.): Privacy and Identity 2022, IFIP AICT 671, pp. 11–20, 2023.
https://doi.org/10.1007/978-3-031-31971-6_2

personal data does not guarantee that this aim is achieved: Several approaches reduce the identifiability of data subjects, but do not yield anonymous data (cf. [7]). Thus, there are multiple questions of interest concerning this theoretical state of anonymity, both from a legal and from a technical side. If data is not "sufficiently anonymised" (i.e. some kinds of anonymisation measures have been applied, but the identifiability of data subjects cannot be excluded to the necessary extent), it would still be considered personal data, hence the GDPR would apply in full—including obligations to protect the data and its processing with appropriate technical and organisational safeguards. Here, privacy-enhancing technologies play a major role, both as safeguards and as data minimisation tools. Concepts like differential privacy and privacy-preserving processing approaches based on e.g. homomorphic encryption or multi-party computation may provide strong guarantees of protection if applied correctly. Still, it is not an automatism that applying such techniques leads to anonymous data or to the level of data protection required by the GDPR. Hence, the major open question here is to determine when an anonymisation technique is "good enough" to reasonably consider its outcome as anonymous. Similarly, for pseudonymisation techniques it would have to be assessed whether the applied techniques result in pseudonymised data (as defined in Article 4(5) GDPR)[1].

Even if an anonymisation or pseudonymisation technique would not, or not always yield anonymised or, respectively, pseudonymised data, it could be valuable or even necessary for fulfilling the demands of the GDPR concerning appropriate technical and organisational measures due to its effects on reducing the risk for rights and freedoms of natural persons. In particular this encompasses technologies for reducing the identifiability of data subjects, e.g. by achieving pseudonymous data.

In this perspective, emerging technologies around machine learning and artificial intelligence play a special role. Machine-learning models need to be trained with input data to fulfil their respective purposes. This training data is often directly linkable to human individuals, therefore clearly not anonymous. Hence, the act of training a model itself may already constitute an act of processing of personal data—with all the legal consequences that arise from the GDPR for this. This poses multiple questions, e.g.: How can reasonable safeguards be set up here? How can they be validated? What level of protection is possible, and which learning approaches substantiate what level of protection of the training data?

Beyond that, also the model itself as outcome of the training phase may be classified as personal data if the linkability to human individuals from the training dataset is maintained by the learning approach. In particular, membership inference attacks [14] have demonstrated that machine-learning models may reveal which data subjects have contributed with their—potentially sensitive—personal data to the model training. For instance, if the model classi-

[1] Note that the GDPR defines the process of "pseudonymisation" with the outcome of "pseudonymised data" which is a subset of all kinds of "pseudonymous data" where the identity of the data subjects is hidden to some extent.

fies a medical disease, the fact that persons contributed data to the model may leak that these persons have this disease. This prompts further questions, e.g.: Under which conditions can a machine-learning model be classified as anonymous or pseudonymous? If the model may still be classified as personal data, under which conditions—potentially including additional technical and organisational measures—would it be lawful to forward the model to other legal entities under the GDPR? Can the linkability to the individuals from the training dataset be removed? Or at least aggregated or hidden to an extent that reasonably well reduces the risk of re-identification to substantiate anonymous data in light of the GDPR? If not, is it at least meeting the demand for strong safeguards with respect to processing?

In this context, recent research on usable privacy emphasises the need to explain privacy-enhancing technologies (PETs) with functional models detailing not only how a PET works but rather "why" it should be used [16], i.e. what are the benefits and implications for users or other types of stakeholders for using a PET. In particular, it has been pointed out that differential privacy should rather be explained as a reduction of the risk of re-identification and their practical implications for users (instead of emphasising other aspects such as privacy-utility trade-offs) [11,12].

Yet another twist in this game is the fact that much of this training of machine-learning models or the use of such models often happens in cloud systems hosted outside of Europe, mostly in the U.S., hence—according to the Schrems II decision of the Court of Justice of the European Union (CJEU)[2] — specific supplementary measures in addition to legal transfer instruments such as Standard Contractual Clauses (SCC) must be taken to legally transfer data to these third countries. Such supplementary measures could include the implementation of strong technical privacy-enhancing safeguards for the processing—which leads to exactly the same set of questions as before.

2 Workshop Summary

In order to address these questions and shed a light on the concept of anonymity in different application scenarios, we organised a one-hour workshop at the IFIP Summer School on Privacy and Identity Management in 2022, held online due to pandemic restrictions. The workshop participants consisted of a broad mix of different backgrounds, ranging from Ph.D. students to senior academics to representatives of industry.

The main task of the workshop consisted in two subsequent exercises around the concept of anonymity. In the first exercise, the participants were asked to align a set of different processing scenarios (with and without naming specific safeguard technologies like homomorphic encryption) along an axis ranging from *not anonymous* via *less anonymous*, *somewhat anonymous*, and *more anonymous* to *truly anonymous*, as is shown in the upper part of Fig. 1 (The lower part of

[2] https://curia.europa.eu/juris/liste.jsf?num=C-311/18.

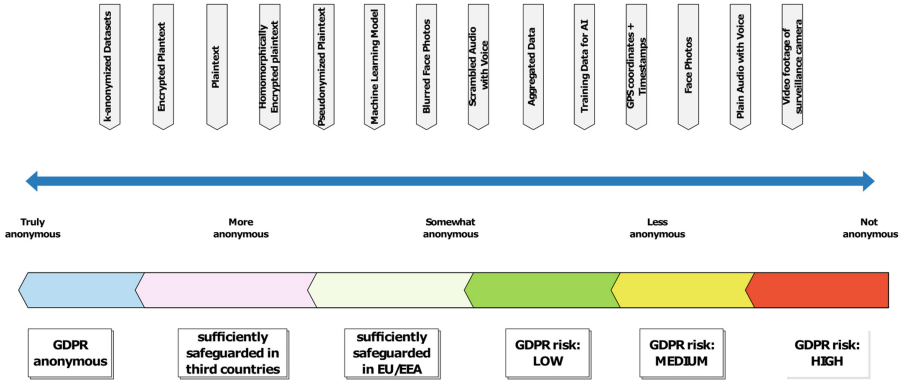

Fig. 1. The conceptboard presented to the workshop participants initially

the figure was hidden to the participants at this stage). This initial arrangement was set up intentionally, in order to foster discussion among the interdisciplinary audience. Obviously, these categories were—on purpose—not aligned with the terminology used in the GDPR or in other approaches for a more sophisticated terminology (cf. [3,13]. In particular, when regarding "anonymity" as a binary concept which directly determines whether the GDPR is applicable or not, there would be no space for "more anonymous" or "less anonymous" or for a notion of different "anonymity levels". A limitation to "truly anonymous" (i.e. "anonymous in the sense of the GDPR") would not have been helpful for fleshing out specific properties with respect to the degree of reducing the risk of re-identification.

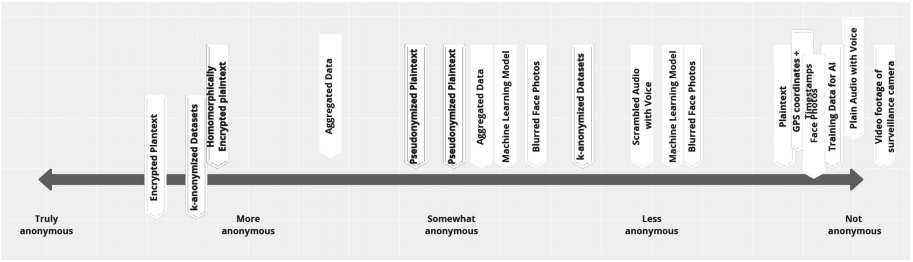

Fig. 2. The conceptboard result of Task 1

As one could expect, the task raised a lot of debates on its feasibility and validity, but led to a predominantly consensual result, where more advanced techniques were considered *more anonymous*, and unprotected data was considered *not anonymous*. No scenario or technique was considered *truly anonymous* (cf. Fig. 2). Along with this, a consensual agreement was that the information

provided per scenario/technique was not sufficient and left a large room for debate and pitfalls, so the consensus was that "it depends". Beyond that, some interesting discussion findings were as follows:

- With the approach of k-anonymity, the participants agreed that the idea is: the higher the k, the *more anonymous* the data.
- With respect to machine-learning models, the more a model is considered explainable, the *less anonymous* the resulting model is,
- The participants agreed that data aggregation is a powerful mechanism, and a kind of slider, for decreasing the identifiability of individuals and thereby supporting anonymity.
- With regards to encrypted plaintext, it depends on who knows the secret key, and that it is only a matter of time until encryption could probably be broken.
- There was no one among the participants who said: something is *truly anonymous*. This was not challenged by anyone.
- There was an intense debate around the effort that is necessary to de-anonymise/re-identify data.
- With respect to "homomorphically encrypted plaintext", it was noted that additional information was kept in the ciphertext for analysis, and that this may leak information. In this line, it was highly debated whether homomorphic encryption was equivalent or weaker than standard symmetric encryption with respect to anonymity protection.
- On risk assessment, the participants stated that even if a risk is not likely, there may be a high damage.
- There was an intense discussion around the concept of emotion detection from video footage. It was stated, but also challenged, whether such a system, when utilising appropriate PETs, would be legal (especially in regard to compliance with the upcoming AI Act that forbids high-risk AI applications) and sufficient with respect to protecting anonymity.

In the second task, the lower part of the conceptboard was revealed, indicating a "mapping" of the given "anonymity levels" to relevant concepts from the data protection law domain. One aspect, taken from the risk assessment approach of data protection impact assessments (cf. Article 35 GDPR), mapped the "anonymity levels" to risk levels of *anonymous* (=no risk), *LOW* risk, *MEDIUM* risk, and *HIGH* risk. Additionally, two more categories shown were *sufficiently safeguarded in EU/EEA* and *sufficiently safeguarded in third countries*, implying that these "levels" were somewhat between *anonymous* and *LOW* risk level. Again, the participants were asked to adjust the position of the scenario/technology markers to this new scale, with results as shown in Fig. 3.

This time, the discussions were more critical concerning the task definitions, as large doubts were raised as to whether it is even possible to map the "anonymity levels" to these categories, and definitely not for the given scenario/technique markers given. The discussion clearly showed that it was not trivially possible to map these different concepts to a scale or to each other

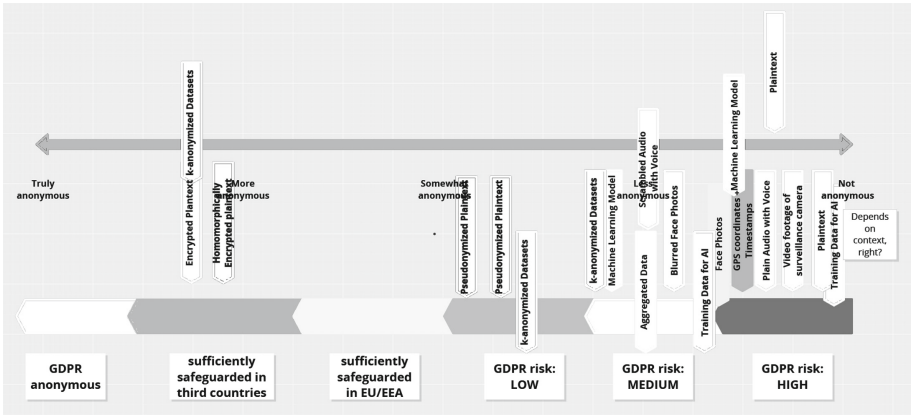

Fig. 3. The conceptboard result of Task 2

as was implied by the task descriptions given, and that more research would definitely be necessary to get to a better understanding of the interrelations of the different concepts interwoven here. Beyond that, some interesting discussion findings were as follows:

- It was suggested to phrase the task question differently: "Suppose I am in a high-impact environment, and the risk is depending on the data: how and how far can I reduce the risk?".
- We wondered on the actual delta of risk reduction, e.g. when data is encrypted compared to non-encrypted. What is the "amount of risk reduction"? Are there some techniques that always reduce the risk? Can we quantify the amount of risk reduction?
- The whole concept of anonymisation was challenged.
- It was discussed that the problem is much bigger: data protection is not the only leverage.
- An interesting find was a scenario in which applying a PET may be worse than not-applying a PET. If the use of PETs in machine learning reduces the accuracy of the trained model (by removing data that was relevant for the model computations), the application of PETs may lead to a situation where the use of the model is no longer "good enough" for the purpose, and should not be used at all.
- We identified a need for transparency, why a certain judgement is taken by a decision algorithm.

3 Open Challenges

In the workshop, we also raised the following questions that still constitute research challenges, as also touched upon above:

When using (data minimisation) PETs in a certain context, under which remaining (residual) risks:

- can data be considered as anonymous under the GDPR?
- may these PETs, potentially together with other measures, be assessed as appropriate technical and organisational measures for complying with the GDPR's principle for data protection by design and by default (Art. 25 GDPR)?
- can these PETs be considered as a supplementary measure for SCC for allowing non-EU Cloud usage (in compliance with the Schrems II CJEU decision and the European Data Protection Board's (EDPB) Recommendation 01/2020, [6])?
- can PETs render high-risk AI systems "acceptable" (in the meaning of compliance with the AI Act)?
- And finally: What other motivations or requirements, from legal, technical, organisational or economical backgrounds, may affect such implementations of PETs in real-world settings?

The objective of posing these questions was to create awareness of challenges faced when approaching the questions (rather then answering them). In the following subsections, we discuss the second and third question further. For the discussion, we consider the use case of a company that plans to perform data analytics of sensitive data (about sick leaves taken by employees). As PETs, we consider data minimisation technologies for data analytics including (local or central) differential privacy.

3.1 Article 25 GDPR

Let us briefly consider Article 25 of the GDPR, entitled "Data protection by design and by default", in a bit more detail. It states in paragraph 1:

> Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.

Although the title of this article appears to suggest that it constitutes an obligation to design systems with data protection as a core design requirement, Bygrave argues [4] the article is quite vague and generic, lacking clear guidelines and incentives to actually "hardwire" privacy-related interests into the design of systems and services. Others, like Jasmontaite et al. [10], do see elements of such an obligation and state that the data controller has to implement both technical and organisational measures in order to ensure that the requirements of the GDPR are effectively embedded in all stages of the processing activity.

Concrete guidelines to support the privacy-friendly design of systems and services from the very start do exist [8,9]. These are supported by additional privacy-enhancing technologies and design approaches further down the design process [5].

The situation is less clear for the specific case of machine learning. The European Data Protection Board, in its Guidelines on Data Protection by Design and by Default [2], note that for automated decision making and artificial intelligence based approaches, *accuracy* is a key concern. In particular because "inaccurate personal data could be a risk to the data subjects' rights and freedoms, for example when leading to a faulty diagnosis or wrongful treatment of a health protocol, or an incorrect image of a person can lead to decisions being made on the wrong basis".

The focus on accuracy is not by accident. The fundamental data protection principle of data minimisation is limited in its effect for machine learning that—by its very nature—requires a lot of detailed information both when being trained and when being used. And supposedly privacy-friendly approaches like federated learning that shift the processing to the end points or end user devices do prevent the *centralised* collection and processing of personal data, but not the processing of personal data per se.

3.2 Third Countries/Cloud Processing

In its judgment C-311/18 (Schrems II), the CJEU clearly pointed out that personal data protection must also be guaranteed if the data is transferred from the European Economic Area (EEA) to a third country. SCC mentioned in Article 46 GDPR were still declared as a valid contractual transfer instrument, but it was emphasised that at the same time SCC need to be complemented by supplementary measures to guarantee a levelof protection of the data transferred up to the EU standard. In Annex 2 of the EDPB's Recommendations 01/2020 [6] on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data, examples of supplementary measures including technical measures are given and discussed, including transfer of pseudonymised data.

While differential privacy or k-anonymity are not specifically listed as examples of technical measures, Stalla-Bourdillon et al. [15] argue that k-anonymity or differential privacy techniques could be considered as sufficiently secure pseudonymisation techniques if they are implemented as privacy-enhancing data transformation measures and sufficiently preclude a risk of re-identification. As pointed out in [14], differentially private models are, by construction, secure against membership inference attacks. Hence, differential privacy applied to machine-learning models (with a sufficient preclusion of the risk of re-identification), could in our use case be regarded as a suitable supplementary measure for outsourcing the model for data analytics e.g. to a non-European cloud service, or for using differential privacy combined with federated learning for creating a central model to be used in the cloud.

Still, given privacy-utility trade-offs that differential privacy implies, challenges remain for achieving a sufficiently low risk that can be accepted for the data processing in the responsibility of the respective controller without compromising on utility and thus on the privacy requirement for data accuracy.

4 Conclusion and Future Research Directions

While our short workshop could by no means solve all problems and answer all questions stemming from the complex situation of machine learning, PETs and the GDPR, the preparation process among the organisers and the interdisciplinary discussion with the participants provided an additional value

- in understanding legal demands from the GDPR and the Schrems II CJEU decision in the field of anonymisation, pseudonymisation and other measures for sufficiently reducing the risks for individuals,
- in comprehending properties, achievements and limitations of specific PETs,
- in grasping challenges concerning specific characteristics of machine learning with respect to personal data in different stages of processing,
- in conceiving the existing difficulties of applying and matching the identified legal demands in the respective field with respect to practical purposes of processing personal data, and
- in fostering a dialogue among researchers interested in privacy and identity management, PET developers, and organisations willing to employ PETs to promote compliance with Article 25 GDPR.

For achieving the objective of clarity on how to apply the GDPR, in particular concerning Article 25 GDPR, and of legal certainty concerning machine learning, manifold research questions have to be tackled in the near future, as described in the previous sections. This encompasses fundamental questions on identifiability as well as best practice solutions on specific cases to bridge the gap between data protection law and practice of development and usage of machine learning.

References

1. Agencia Española de Protección de Datos and European Data Protection Supervisor. 10 misunderstandings related to anonymization (2021). https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf
2. European Data Protection Board. Guidelines 4/2019 on article 25. data protection by design and by default (2020). https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-42019-article-25-data-protection-design-and_en
3. Bruegger, B.P.: Towards a better understanding of identification, pseudonymization, and anonymization (2021). https://uld-sh.de/PseudoAnon
4. Bygrave, L.: Data protection by design and by default: deciphering the EU's legislative requirements. Oslo Law Rev. **4**(2), 105–120 (2017)
5. Danezis, G., et al.: Privacy and Data Protection by Design - from policy to engineering. Technical report, ENISA (2014). ISBN 978-92-9204-108-3. https://doi.org/10.2824/38623. https://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/privacy-and-data-protection-by-design

6. European Data Protection Board. Recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data (2020). https://edpb.europa.eu/system/files/2021-06/edpb_recommendations_202001vo.2.0_supplementarymeasurestransferstools_en.pdf
7. Finck, M., Pallas, F.: They who must not be identified-distinguishing personal from non-personal data under the GDPR. Int. Data Privacy Law **10**(1), 11–36 (2020). https://doi.org/10.1093/idpl/ipz026. ISSN 2044-3994
8. Hoepman, J.-H.: Privacy design strategies. In: Cuppens-Boulahia, N., Cuppens, F., Jajodia, S., Abou El Kalam, A., Sans, T. (eds.) SEC 2014. IAICT, vol. 428, pp. 446–459. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-55415-5_38
9. Hoepman, J.-H.: Privacy design strategies. The little blue book (2018). https://www.cs.ru.nl/jhh/publications/pds-booklet.pdf
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. Eur. Data Prot. Law Rev. **4**(2), 168–189 (2018)
11. Karegar, F., Alaqra, A.S., Fischer-Hübner, S.: Exploring user-suitable metaphors for differentially private data analyses. In: Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022), Boston, MA, pp. 175–193. USENIX Association (2022). https://www.usenix.org/conference/soups2022/presentation/karegar. ISBN 978-1-939133-30-4
12. Nanayakkara, P., Bater, J., He, X., Hullman, J., Rogers, J.: Visualizing privacy-utility trade-offs in differentially private data releases. arXiv preprint arXiv:2201.05964 (2022)
13. Pfitzmann, A., Hansen, M.: A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management, vol. 34 (2010). http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf
14. Shokri, R., Stronati, M., Song, C., Shmatikov, V.: Membership inference attacks against machine learning models. In: 2017 IEEE Symposium on Security and Privacy (SP), pp. 3–18. IEEE (2017)
15. Stalla-Bourdillon, S., Rossi, A.: Why a good additional technical safeguard is hard to find–a response to the consultation on the EDPB draft recommendations 01/2020 on measures that supplement transfer tools to ensure compliance with the EU level of protection of personal data (2020). https://edpb.europa.eu/sites/default/files/webform/public_consultation_reply/response_sto_edpb_recommendations.pdf
16. Wu, J., Zappala, D.: When is a tree really a truck? Exploring mental models of encryption. In: Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018), pp. 395–409 (2018)