Bridging the gap between technology and policy in GDPR compliance: the role of differential privacy*

Asmaa Belghiti[1], Armando Angrisani[2]

[1]*NEOMA Business School, Paris, France*
[2]*LIP6, CNRS, Sorbonne Université, Paris, France,*

Participation in our digital society requires the continuous sharing of sensitive information, enabling the implementation of sophisticated machine learning algorithms. Along with technological benefits, these societal transformations come with undesired effects, namely the erosion of individual privacy. So far, policy-makers have failed to account for the potential harms of extensive data collection and processing. The adoption of the EU General Data Protection Regulation (GDPR)[1] partially addressed this issue, granting individuals the right to erase their data held by firms. However, GDPR also allows firms to retain aggregate statistics, provided that they are sufficiently "anonymized", following Recital 26. Such "anonymization" is addressed with various techniques, ranging from heuristics, such as *pseudonymization* (Article 4), to more rigorous approaches, such as *differential privacy*[2;3] and *k-anonymity*[4]. A recent line of research consists in translating in mathematical terms the prescriptions of the GDPR and comparing them with existing privacy-preserving tools. In particular, the mathematical notion of *PSO security* formalizes the concept of "singling out", introduced in Recital 26[5]. This connects the GDPR with the techniques mentioned above: differential privacy provides PSO security, while k-anonymity and pseudonymization do not. However, the implementation of differential privacy involves several caveats and subtleties. A DP algorithm comes equipped with a parameter ε, which measures the "level" of privacy. Low values of ε are necessary to ensure meaningful privacy guarantees, but they usually lead to a loss of accuracy. This drawback is something referred to as the *privacy-utility trade-off*[6;7] and it is particularly concerning for the analysis of microdata records[8]. For this reason, many practitioners set the value of ε excessively large, leading to a form of *privacy-washing*[9].

The inherent indeterminacy of GDPR, along with the lack of scientific consensus about privacy, poses several open questions for researchers and policy-makers[10;11]. On the one hand, it is still unclear which privacy-preserving techniques meet the existing legal standards, thus we complement the previous investigation[5] by comparing the *right to erasure* (Article 17) with the novel algorithmic framework of *machine unlearning*[12]. Moreover, we examine possible approaches to incorporate the capabilities and limitations of modern techniques into the current legislation. On the other hand, we review the issues encountered in several industrial implementations of differential privacy[13;14]. We remark that the ε values are often omitted, with concrete risk of privacy-washing. Hence, we claim that firms should make public the details of their implementations[15]. This would enable stakeholders to compare the quality of privacy offered by various firms and create pressure to reduce privacy losses. Moreover, an adequate compensation policy for infringement of the GDPR (Article 82) could effectively contrast privacy-washing, enforcing the value of ε indirectly. Indeed, the privacy-utility trade-off and the compensation policy determine an optimal value of ε that minimizes the expected economic loss of firms[16]. Finally, we advocate for educational initiatives aiming at the diffusion of "privacy literacy". While a complete understanding of the computational issues requires technical training, the large public could grasp the main concepts of differential privacy through interactive simulations and intuitive user interfaces[9;17;18].

*Contributed talk at "Privacy 2.0 - Interdisciplinary perspectives on privacy in the digital age", Conference of the Hans Böckler Foundation, 7 April 2022, Dresden, Germany.

References

[1] General Data Protection Regulation. Regulation (EU) 2016/679 of the European Parliament and of the European Council of 27 April 2016.

[2] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, pages 265–284, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-32732-5. URL https://link.springer.com/chapter/10.1007/11681878_14.

[3] Rachel Cummings and Deven Desai. The role of differential privacy in GDPR compliance. 2018. URL https://cpb-us-w2.wpmucdn.com/sites.gatech.edu/dist/c/679/files/2018/09/GDPR_DiffPrivacy.pdf.

[4] Pierangela Samarati and Latanya Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, 1998. URL http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.37.5829.

[5] Aloni Cohen and Kobbi Nissim. Towards formalizing the GDPR's notion of singling out. *Proceedings of the National Academy of Sciences*, 117(15):8344–8352, Mar 2020. ISSN 1091-6490. doi: 10.1073/pnas.1914598117. URL http://dx.doi.org/10.1073/pnas.1914598117.

[6] Mário S. Alvim, Miguel E. Andrés, Konstantinos Chatzikokolakis, Pierpaolo Degano, and Catuscia Palamidessi. Differential privacy: On the trade-off between utility and information leakage. *Formal Aspects of Security and Trust*, page 39–54, 2012. ISSN 1611-3349. doi: 10.1007/978-3-642-29420-4_3. URL http://dx.doi.org/10.1007/978-3-642-29420-4_3.

[7] Sushant Agarwal. Trade-offs between fairness and privacy in machine learning. In *IJCAI 2021 Workshop on AI for Social Good*, 2021. URL https://crcs.seas.harvard.edu/publications/trade-offs-between-fairness-and-privacy-machine-learning.

[8] Steven Ruggles, Catherine Fitch, Diana Magnuson, and Jonathan Schroeder. Differential privacy and census data: Implications for social and economic research. *AEA Papers and Proceedings*, 109:403–08, May 2019. doi: 10.1257/pandp.20191107. URL https://www.aeaweb.org/articles?id=10.1257/pandp.20191107.

[9] Cynthia Dwork, Nitin Kohli, and Deirdre Mulligan. Differential privacy in practice: Expose your epsilons! *Journal of Privacy and Confidentiality*, 9(2), Oct. 2019. doi: 10.29012/jpc.689. URL https://journalprivacyconfidentiality.org/index.php/jpc/article/view/689.

[10] Vitaly Feldman, Konstantin Kakaes, Katrina Ligett, Kobbi Nissim, Aleksandra Slavkovic, and Adam Smith. Differential privacy: Issues for policymakers, Aug 2020. URL https://simons.berkeley.edu/news/differential-privacy-issues-policymakers.

[11] Daniel L. Oberski and Frauke Kreuter. Differential privacy and social science: An urgent puzzle. *Harvard Data Science Review*, 2(1), 1 2020. doi: 10.1162/99608f92.63a22079. URL https://hdsr.mitpress.mit.edu/pub/g9o4z8au.

[12] Ayush Sekhari, Jayadev Acharya, Gautam Kamath, and Ananda Theertha Suresh. Remember what you want to forget: Algorithms for machine unlearning, Dec 2021. URL https://proceedings.neurips.cc/paper/2021/hash/9627c45df543c816a3ddf2d8ea686a99-Abstract.html.

[13] Solomon Messing, Christina DeGregorio, Bennett Hillenbrand, Gary King, Saurav Mahanti, Zagreb Mukerjee, Chaya Nayak, Nate Persily, Bogdan State, and Arjun Wilkins. Facebook Privacy-Protected Full URLs Data Set, 2020. URL https://doi.org/10.7910/DVN/TDOAPG.

[14] Jun Tang, Aleksandra Korolova, Xiaolong Bai, Xueqiang Wang, and Xiaofeng Wang. Privacy loss in apple's implementation of differential privacy on macos 10.12, 2017. URL https://arxiv.org/abs/1709.02753.

[15] Cynthia Dwork, Nitin Kohli, and Deirdre Mulligan. Differential privacy in practice: Expose your epsilons! *Journal of Privacy and Confidentiality*, 9(2), Oct. 2019. doi: 10.29012/jpc.689. URL https://journalprivacyconfidentiality.org/index.php/jpc/article/view/689.

[16] Ashish Dandekar, Debabrota Basu, and Stéphane Bressan. Differential Privacy at Risk: Bridging Randomness and Privacy Budget. *Proceedings on Privacy Enhancing Technologies*, 2021(1):64–84, 2021. doi: 10.2478/popets-2021-0005. URL https://hal.inria.fr/hal-02942997.

[17] Rachel Cummings, Gabriel Kaptchuk, and Elissa M. Redmiles. "I need a better description": An investigation into user expectations for differential privacy. *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021. URL https://dl.acm.org/doi/10.1145/3460120.3485252.

[18] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605587363. doi: 10.1145/1572532.1572538. URL https://doi.org/10.1145/1572532.1572538.