Introducing Red Teaming for Research Data Anonymization

Luisa Jansen
University of Bern, Switzerland

Tim Ulmann
University of Bern, Switzerland

Robine Jordi
University of Bern, Switzerland

Malte Elson
University of Bern, Switzerland

Abstract

Recently, the data protection practices of researchers in Usable Security and Privacy and elsewhere have gained attention, with initial results suggesting that researchers struggle with anonymization, partly due to a lack of clear, actionable guidance. We propose simulating re-identification attacks using red teaming, a technique from security testing in which one team attempts to re-identify data while another tries to prevent it. We report our experience applying this method to data from a mixed-methods study in human-centered privacy, present usable materials to apply red teaming, and discuss limitations, ethical implications, and directions for future work.

1 Introduction

Recent work has shed light on research practices that usually remain hidden: the data protection practices of empirical research data [4, 9], indicating researchers' insecurities and struggles surrounding data anonymization. With open data gaining more momentum in Usable Security and Privacy (USP) research, anonymization becomes more important [5, 10]. The commit to keeping personal data private is enforced by both ethical commitments within the research community [2, 13] and many data protection laws worldwide (e.g., the General Data Protection Regulation [11]).

When seeking advice on anonymizing research data before sharing, researchers are likely to encounter the ubiquitous saying "as open as possible, as closed as necessary" [3, 7]. What this means in practice remains obscure. Well-known privacy concepts such as k-anonymity are often recommended for research data, but their application relies on a series of judgment calls at the threat modeling stage, which are especially uncertain given the complexity of research projects [4].

We therefore propose a procedure to navigate the trade-off between protecting and opening data: We apply the concept of red teaming to determine the actual risks associated with data sharing by attacking our own anonymization[1]. Red teaming has its roots in security practice [1] and is nowadays a common method for improving the cybersecurity of organizations [8, 12]. The process is typically iterative, meaning that the red and blue teams respond to each other's actions in multiple rounds or continuously over a specified time period [8]. Applying this method to the anonymization of research data, we hope to contribute to the privacy efforts of researchers in USP and elsewhere by making common privacy advice more usable and actionable. We provide materials for researchers to apply the method themselves and demonstrate red teaming for data anonymization with one example from our own research.

2 The Method

The process begins with a minimally anonymized research dataset. In the first step, the red team plays the role of the attacker: They attempt to re-identify research participants using any publicly available information about the study (e.g., recruitment strategy) alongside the data itself. The red team is free to follow any path and can therefore be very creative. They report their results to the blue team. In the second step, the blue team uses the red team's findings to improve the anonymization by using anonymization techniques. Blue team members follow a more structured approach and ideally have in-depth knowledge of the materials to be published. They must continuously balance privacy risks against data utility. These two steps form one iteration. The newly anonymized data, as well as potentially revised study materials, are then subjected to another round of red teaming. If the red team

USENIX Symposium on Usable Privacy and Security (SOUPS) 2026.
August 23–26, 2026, Hannover, Germany.

[1] A longer preprint and a practical guide for using the method are accessible at: https://doi.org/10.48550/arXiv.2601.19575

succeeds in re-identifying participants, the blue team can further strengthen the anonymization. The process continues until the teams agree the risk is sufficiently reduced.

3 Case Report: Applying Red Teaming

We applied the method ourselves to the data associated with a mixed-method expert study conducted by Luisa Jansen (A1) and others (link to study and data) [6]. A1 acted as the blue team, while Tim Ulmann (A2) and Robine Jordi (A3) acted as the red team.

The **underlying study** was a mixed-methods expert study in human-centered privacy. 12 privacy experts were recruited using a freelancing platform. The study evaluated an intervention designed to improve the communication skills of privacy expert within two groups. The data included demographics, ratings of the intervention, and solutions to communication tasks. When publishing the study, the aim was to share the data as openly as possible. **Before starting** the process, A1, acting as the blue team, minimally anonymized the data by replacing personal codewords with random numbers.

Iteration 1 – Red team: A3 started by reverse-engineering the recruitment search strategy on the freelancing platform used in the study and filtering for participants with uncommon demographic characteristics. This led to the identification of profiles for four participants, including their full names. Their participation could be verified through publicly available profile information. Using this information, A3 was able to link these individuals to all other study data.

Iteration 1 – Blue team: A1 focused on mitigating the most effective attack vector: identifying participants via their country of residence. To address this, countries were recoded into meaningful categories. In addition, demographics were de-associated from all other study data to prevent linking individual demographics to performance. Demographic information was further aggregated by reporting summaries for each study group rather than individual-level demographics. These changes did not meaningfully reduce data utility.

Iteration 2 – Red team: A2 took over the red team and again used the freelancing platform to reverse-engineer the recruitment strategy. A2 found overlooked information in the accompanying materials where participants' countries of residence were still visible, though without allowing linkage to the other study data. Over approximately eight hours, A2 was able to find the identities of two participants. Their participation could be verified through references to the study found in their public profiles. A2 also attempted to use Qualtrics participation timestamps to estimate participants' time zones and tried to match linguistic styles in the study data to public freelancing profiles. Overall, A2 could only form tentative inferences.

Iteration 2 – Blue team: A1 suppressed dates and times of participation and removed information about participants' countries of residence from the accompanying materials.

Ethics. Throughout, all work was conducted in compliance with the General Data Protection Regulation (GDPR): red and blue team members were part of the same research group, and any access to personal data was covered by informed consent. The mixed-methods study was approved by our ethics committee. In our assessment, red teaming does not introduce additional risks, but rather serves as a controlled process to identify vulnerabilities before they can be exploited by external actors.

4 Discussion and Outlook

Our approach has several limitations. First, red teaming does not eliminate all risks. It does not systematically uncover all possible attack vectors, and its effectiveness depends on the creativity, expertise, and persistence of the team members. The process can also be resource-intensive, requiring additional time and effort from the research team.

Ethical aspects also require further discussion. In particular, one may question whether the potential risk of re-identification by a broader public justifies controlled attempts by a red team to identify participants' identities. In this context, approaches that go beyond traditional one-time consent models and involve participants more directly in anonymization decisions deserve further consideration. Ideally, participants should at least be informed about and give consent to potential red-teaming procedures as part of the study's consent process.

Despite these limitations, the approach offers several advantages. While the available resources influence the depth of testing, the red-teaming approach provides a flexible framework that can be adapted to different research contexts and resource levels. Even when not all attack vectors can be explored, the method allows researchers to identify at least the most obvious vulnerabilities and thereby reduce the likelihood of successful re-identification with reasonable effort. In this sense, the approach offers a practical way to operationalize the principle of making data "as open as possible, as closed as necessary." From both legal and ethical perspectives, we therefore consider red teaming a useful tool to support responsible data publication. Finally, although anonymization is often seen as a rather technical or "dry" topic, we perceived the red-teaming process as engaging and enjoyable.

Looking ahead, we plan further testing to better understand for which types of studies this approach is most suitable and whether researchers are willing and able to adopt it in practice. We invite discussions to explore these questions, gather feedback, and refine the method collaboratively.

References

[1] Hussein Abbass, Axel Bender, Svetoslav Gaidow, and Paul Whitbread. Computational red teaming: Past,

present and future. *IEEE Computational Intelligence Magazine*, 6(1):30–42, 2011.

[2] Jacob Abbott, Haley MacLeod, Novia Nurain, Gustave Ekobe, and Sameer Patil. Local standards for anonymization practices in health, wellness, accessibility, and aging research at CHI. In *CHI Conference on Human Factors in Computing Systems (CHI '19)*, pages 1–14, Glasgow Scotland Uk, May 2019. ACM.

[3] European Research Executive Agency. Open Science. https://rea.ec.europa.eu/open-science_en.

[4] Wentao Guo, Paige Pepitone, Adam J Aviv, and Michelle L Mazurek. How researchers de-identify data in practice. In *Proceedings of the 34th USENIX Security Symposium*, Seattle, WA, USA, 2025.

[5] Luisa Jansen, Nele Borgert, and Malte Elson. On the tension between open data and data protection in research, April 2025.

[6] Luisa Jansen, Stefan Albert Horstmann, Nele Borgert, Alena Naiakshina, and Malte Elson. Bridging the gap: A training to enhance privacy experts' communication with software developers through audience design. 2023.

[7] Annalisa Landi, Mark Thompson, Viviana Giannuzzi, Fedele Bonifazi, Ignasi Labastida, Luiz Olavo Bonino Da Silva Santos, and Marco Roos. The "A" of FAIR – as open as possible, as closed as necessary. *Data Intelligence*, 2(1-2):47–55, 2020.

[8] Steve Mansfield-Devine. The best form of defence: The benefits of red teaming. *Computer Fraud & Security*, 2018(10):8–12, 2018.

[9] Florin Martius, Luisa Jansen, Lukas Struck, Arthi Arumugam, Lisa Geierhaas, Anna-Marie Ortloff, Matthew Smith, and Christian Tiefenau. Out of sight, out of mind? Exploring data protection practices for personal data in usable security & privacy studies. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, Yokohama, Japan, 2025. ACM.

[10] Jonas Oppenlaender, Sylvain Malacria, Xinrui Fang, Niels van Berkel, Fanny Chevalier, Koji Yatani, and Simo Hosio. Meta-HCI: First workshop on meta-research in HCI. *CHI EA '25, Yokohama, Japan*, 2025.

[11] European Parliament and Council of the European Union. General Data Protection Regulation, April 2016.

[12] Johann Rehberger. *Cybersecurity Attacks - Red Team Strategies: A Practical Guide to Building a Penetration Testing Program Having Homefield Advantage*. Packt Publishing, 1st edition, 2020.

[13] Kavous Salehzadeh Niksirat, Lahari Goswami, Pooja S. B. Rao, James Tyler, Alessandro Silacci, Sadiq Aliyu, Annika Aebli, Chat Wacharamanotham, and Mauro Cherubini. Changes in research ethics, openness, and transparency in empirical studies between CHI 2017 and CHI 2022. In *CHI Conference on Human Factors in Computing Systems (CHI '23)*, pages 1–23, Hamburg Germany, April 2023. ACM.