Exploring the Use of Interactive Interfaces and Feedback Mechanisms to Enhance Privacy in Data Workers through Information Accountability

Hye-Chung Kum, *Texas A&M University* Eric D. Ragan, *University of Florida*

Abstract

While most privacy and security solutions aim to entirely prevent access or disclosure of sensitive data, many types of data workers, researchers, and developers commonly require access to sensitive data as part of their jobs. For instance, medical and social science research requires analysis of personal information to draw conclusions, inspect anomalies, and verify data quality [1]. Or consider database administration, which often requires reviewing data composition and quality to make decisions about database designs or for data cleaning [2]. As another example, security analysts investigate detailed data content and transmissions as part of their analyses [3]. Even for the software design of data management systems and interfaces, designers often consider real scenarios and data items as part of requirements gathering and use case analyses.

The problem is that for such types of data work, the inability to access the underlying data details can negatively influence decision making and data utility; pure encryption or data hiding are not sufficient to support good evidence-based decision making [4], [5]. On the other hand, unbridled data access obviously create terrible privacy risks. The reality is that many data workers and intelligence analysts do operate in an all-or-nothing fashion for a given data set. Researchers, developers, and administrators often suffer significant challenges or delays due to access permissions for legitimate data work, or workers have free access to data sets at their own discretion. To add further complication, real-world communities often involve situations where data workers might have access to sensitive organizational or personal information to specific people in the local community (e.g., a database administrator might end up learning sensitive information about friends, co-workers, or neighbors), and practical work environments involve elements of informal information sharing [6].

We contend that there may be no perfect solution to privacy in cases where security workers require some access and knowledge of sensitive information for legitimate and effective utility for many practical purposes. However, we posit that accountability and transparency maybe effective in finding the right balance between access to all data or no data. Furthermore, we suggest that there is a need to explore software visualization designs that encourage ethical behavior and consideration for privacy-aware data practices in practical data work environments.

Due to the significant degree of human decision-making for data access and social privacy, privacy support for data work inherently is largely a human-oriented challenge. There is a critical need for attention to user interface design considerations that allow sufficient flexibility in data work while maintaining accountability through transparent logging, discouraging unnecessary information access, and building user interfaces that encourage privacy-awareness.

We posit that such methods can align with legal requirements for privacy. One of the main principles central to many data protection laws is the *minimum necessary* or *need-to-know* information disclosure standards. Laws like the *Health Insurance Portability and Accountability Act* (HIPAA), the *Privacy Act of 1974*, and the confidentiality protections for substance abuse disorder records in *42 CFR Part 2* use similar legal standards to permit legitimate uses of data while protecting privacy by limiting extraneous disclosures. Similarly, the EU General Data Protection Regulation (GDPR), uses the principle of "data minimization" to limit data use to what is necessary for a permitted purpose.

Thus, we propose to discuss a new class of interface design considerations as a potential approach for balancing tradeoffs between privacy and utility of sensitive information in data work. Interface functionality that allows interactive access only to necessary specific items can limit data disclosure while logging access requests to enable accountability, and appropriate feedback to the user can discourage unnecessary access and promote privacy awareness. In addition, interactive interfaces that support just-in-time decision to access data at the moment of decision making may significantly reduce the need to access data.

To date, we have explored and studied such interface solutions in the context of data cleaning and record linking applications [7], [8]. Our findings have indicated the potential for reducing privacy risks in real settings. In addition, we will discuss general implications when these techniques are applied to dashboard systems used by data workers. Finally, we consider applications for network security analysis tools.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2019. August 11 -- 13, 2019, Santa Clara, CA, USA.

References

[1] H.-C. Kum, A. Krishnamurthy, A. Machanavajjhala, and S. C. Ahalt, "Social genome: Putting big data to work for population informatics," *Computer*, vol. 47, no. 1, pp. 56–63, 2013.

[2] S. Kandel *et al.*, "Research directions in data wrangling: Visualizations and transformations for usable and credible data," *Information Visualization*, vol. 10, no. 4, pp. 271–288, 2011.

[3] J. R. Goodall *et al.*, "Situ: Identifying and explaining suspicious behavior in networks," *IEEE transactions on visualization and computer graphics*, vol. 25, no. 1, pp. 204–214, 2019.

[4] I. Dinur and K. Nissim, "Revealing information while preserving privacy," in *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 2003, pp. 202–210.

[5] S. E. Fienberg, "Confidentiality and disclosure limitation," *Encyclopedia of Social Measurement*, vol. 1, pp. 463–69, 2005.

[6] E. C. O'Brien *et al.*, "Patient perspectives on the linkage of health data for research: Insights from an online patient community questionnaire," *International Journal of Medical Informatics*, vol. 127, pp. 9–17, 2019.

[7] E. D. Ragan, H.-C. Kum, G. Ilangovan, and H. Wang, "Balancing Privacy and Information Disclosure in Interactive Record Linkage with Visual Masking," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018, p. 326.

[8] H.-C. Kum, A. Krishnamurthy, A. Machanavajjhala, M. K. Reiter, and S. Ahalt, "Privacy preserving interactive record linkage (PPIRL)," *Journal of the American Medical Informatics Association*, vol. 21, no. 2, pp. 212–220, 2014.