THE BRUSSELS PRIVACY SYMPOSIUM CALL FOR PAPERS //

Deidentification—broadly understood as the process of modifying personal data to ensure that data subjects are no longer identifiable—is one of the primary measures that organizations use to protect privacy. The reason is simple: in both the EU and the US, privacy and data protection laws do not apply, or apply only in part, to non-identifiable data. Thus, scientific institutions that regularly process, transfer or release large data sets for research purposes rely extensively on deidentification techniques to ensure regulatory compliance.

Similarly, commercial firms that provide financial, healthcare, retail or marketing services often rely on de-identified data for analysis, product improvement and product development. The new EU General Data Protection Regulation (GDPR) introduces the related concept of "pseudonymization," defined as the processing of personal data in such a way as to prevent attribution to an identified or identifiable person without additional information that is held separately. Although pseudonymous data remains subject to the remit of the Regulation, it reduces the risks for data subjects. Consequently, the GDPR relaxes certain requirements on controllers that use the technique for research and statistical purposes, and may allow pseudonymization to be a factor when considering the compatibility of different uses of data. The GDPR also states that the principles of data protection should not apply to anonymous information.

In recent years, several well-publicized incidents have shown that data sets that have apparently been deidentified remain vulnerable to reidentification attacks.[1, 2] These incidents have raised serious doubts for many about the extent to which deidentification remains a credible method for using and deriving value from large data sets while protecting privacy. Both legal and technical experts are sharply divided on the efficacy of deidentification and related solutions. Some critics argue that it is impossible to eliminate privacy harms from publicly released data using deidentification because other available data sets will allow attackers to identify data subjects through linkage attacks.[3, 4] Defenders of deidentification counter that despite the theoretical and demonstrated ability to mount such attacks, the likelihood of reidentification for most data sets remains minimal. As a practical matter, they argue most data sets remain securely deidentified based on established techniques.[5] A similar debate plays out in the technical literature between, on the one hand, researchers who value practical solutions for sharing useful data to advance the public good and therefore devise methods for measuring and managing the risk of reidentification in clinical trials and other research scenarios[6], and, on the other hand, computer scientists seeking mathematical rigor in defining privacy, modeling adversaries, and quantifying the possibility of reidentification.[7] These debates have led some commentators to advocate a new approach in which organizations assess their risk and tailor their obligations accordingly, relying on the full spectrum of technical, contractual and statutory protections against reidentification.[8, 9]

The deidentification debate also overlaps with discussions about "open data." Adherents of an open data philosophy typically support greater access to government (and even corporate) data sets to advance the public good.[10] A key argument in favor of open data within the scientific community is that openness promotes transparency, reproducibility, and more rapid advancement of new knowledge and discovery. Indeed, many scientific journals and funding agencies now require that experimental data is made publicly available; however, they remain divided over what steps researchers must take to protect individuals' privacy before releasing data sets in the open. Making data that have been collected by governments and corporate actors openly accessible can bring data protection and privacy risks, since such data may be highly sensitive. In addition, individuals may have had little choice to provide the data and may not be aware that such data may one day become widely distributed (or even public) and used for secondary purposes. In short, deidentification plays a central role in current privacy policy, law and

practice, notwithstanding the lack of consensus over how best to advance the discussion. The use of open data holds great promise, but also brings risk. And yet the need for sound principles governing data release has never been greater.

To address these challenges, the *Brussels Privacy Symposium*, which is a joint program of the Brussels Privacy Hub of the Vrije Universiteit Brussel (Free University of Brussels or VUB) and the Future of Privacy Forum (FPF) is hosting an academic workshop on *Deidentification: Practical Solutions for Preserving the Social Utility of Data*. **Authors from multiple disciplines including law, computer science, statistics, engineering, social science, ethics and business are invited to submit papers for presentation at a full-day program to take place in Brussels on November 7/8, 2016.** Successful submissions may address issues such as the following:

- **Technology.** Which existing tools or scientific techniques support privacy protective use of datasets by researchers? Is there a conflict between the needs of researchers and existing deidentification standards? How granular is data that is legitimately needed by researchers? What is the current state of the art in technological methods and tools for ensuring safe data release? How do these methods and tools balance competing requirements such as privacy, utility, and efficiency? What are the limitations of different principles and techniques? Are there specific research fields, research questions, or types of data that certain tools are better suited for than others? How practically applicable and scalable are state of the art theoretical solutions such as differential privacy and homomorphic encryption?
- **Policy.** What are the core elements of a data release policy (e.g., consent, data use restrictions, security, accountability)? Are there optimal ways to combine these elements? Are there examples of highly successful projects in which data release successfully balances privacy, utility, and efficiency? Are there best practices that can be derived from such successful projects?
- **Regulation.** What lessons can be learned from existing regulatory mechanisms? How does the concept of "singling out" fit into technical deidentification policy? What are the strengths and weaknesses of generally applicable guidelines such as the anonymization code of practice issued by the United Kingdom's Information Commissioner's Office (ICO) or the opinion on Anonymisation Techniques of the Article 29 Data Protection Working Party compared to sectoral models such as the "Safe Harbor" method for deidentifying health information under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) or the final rule on genomic data sharing issued by the National Institutes of Health (NIH)? Which elements of soft law drawn from the interpretations of data protection authorities around de-identification will continue to apply under the GDPR?
- **Ethics.** How should privacy risks inherent in deidentified data be measured against the potential benefits of data research? How should deidentification standards interact with additional requirements for data research including informed consent by data subjects and review by ethical boards?
- **Open data.** What are the key principles of open data and when is broad dissemination necessary for scientific research and innovation? Should open data rely on technological, policy, or legal tools to protect the privacy interests of data subjects or some combination thereof? Alternatively, is it possible to achieve many of the benefits of open access to data without unrestricted release of data to the public?
- **Pseudonymization.** What technical and organizational measures are required under the GDPR to satisfy the notion of pseudonymization? When organizations utilize such measures, which legal requirements are relaxed under the GDPR? Does this treatment of pseudonymized data provide sufficient incentives for organizations to use this technique as part of an overall compliance strategy?
- **New approaches.** Should privacy policy adopt a new approach to the problems associated with deidentification by focusing less on the ultimate goal of anonymization and more on the processes necessary to lower the risk of reidentification and sensitive attribute disclosure?

An academic advisory board will choose papers for presentation at the workshop. Selected papers will be considered for publication in a special symposium of International Data Privacy Law, a law journal published by Oxford University Press (subject to the journal's normal editorial procedures).

Submissions must be 2,500 to 3,500 words with minimal footnotes and in a readable style accessible to a wide academic audience. Submissions must be made no later than August 1, 2016, at 11:59 PM ET, to **papersubmissions@fpf.org**. Publication decisions and workshop invitations will be sent in September.

---

1 Arvind Narayanan & Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets*, 2008 PROC. 29TH IEEE SYMP. ON SECURITY & PRIVACY 111.

2 Yaniv Erlich & Arvind Narayanan, *Routes for Breaching and Protecting Genetic Privacy*, 15 GENETICS 409 (2014).

3 Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. Rev. 1701 (2010).

4 Arvind Narayanan & Vitaly Shmatikov, *Myths and Fallacies of "Personally Identifiable Information*," 53 COMM. OF THE ACM 24, 26 (2010).

5 Jane Yakowitz, Tragedy of the Data Commons, 25 HARV. J.L. & TECH.1, 2-3. (2011); Ann Cavoukian & Khaled El Emam, *Dispelling the Myths Surrounding Deidentification: Anonymization Remains a Strong Tool for Protecting Privacy* (2011), **http://www.ipc.on.ca/images/Resources/anonymization.pdf**.

6 Khaled El Emam and Bradley Malin, "Appendix B: Concepts and Methods for Deidentifying Clinical Trial Data," in INSTITUTE OF MEDICINE (IOM), SHARING CLINICAL TRIAL DATA: MAXIMIZING BENEFITS, MINIMIZING RISK 7 (2015).

7 Arvind Narayanan & Edward Felten, *No Silver Bullet: De-identification Still Doesn't Work*, http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf.

8 Ira Rubinstein & Woodrow Hartzog, Anonymization and Risk, WASH. L. REV. (forthcoming 2016).

9 Jules Polonetsky, Omer Tene & Kelsey Finch, Shades of Gray: Seeing the Full Spectrum of Practical Data De-Identification, SANTA CLARA L. REV. (forthcoming 2016).

10 The Open Data Institute, "What is Open Data," http://theodi.org/what-is-open-data.