BIG DATA

# The big data debate

By **Fred H. Cate**

"Big data"—the collection, aggregation or federation, and analysis of vast amounts of increasingly granular data—present serious challenges not only to personal privacy but also to the tools we use to protect it. *Privacy, Big Data, and the Public Good* focuses valuable attention on two of these tools: notice and consent, and de-identification—the process of preventing a person's identity from being linked to specific data. The book presents a collection of essays from a variety of perspectives, in chapters by some of the heavy hitters in the privacy debate, who make a convincing case that the current framework for dealing with consumer privacy does not adequately address issues posed by big data.

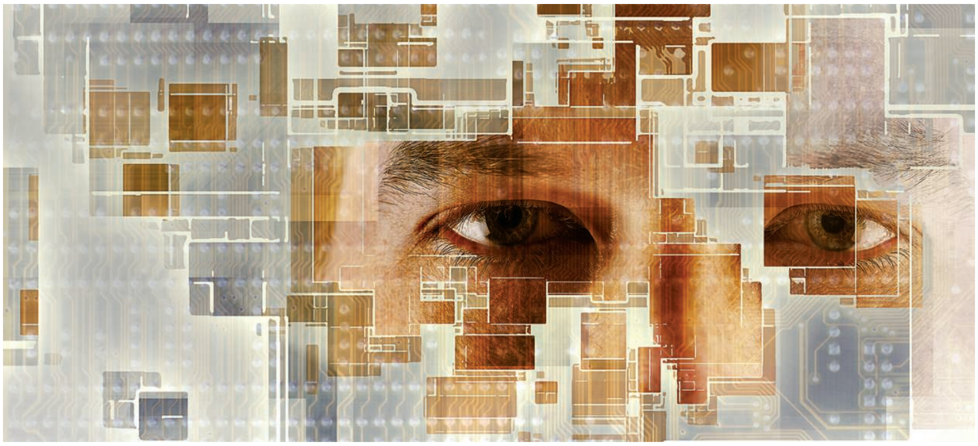

PHOTO: CHAD BAKER AND RYAN MCVAY/THINKSTOCK

In 1998, the U.S. Federal Trade Commission identified notice and consent as the "most fundamental" principles for protecting privacy (*1*). In 2012, the U.S. Consumer Privacy Bill of Rights included as its first principle: "Consumers have a right to exercise control over what personal data companies collect from them and how they use it" (*2*). The pending draft of the European Union Data Protection Regulation refers to "consent" more than 100 times (*3*). Regulators, it seems, are entranced by notice and consent. It is perhaps not surprising, then, that we are inundated with privacy notices and demands that we click "I agree" when we visit banks, doctors, Web sites, app stores, and the like. Yet, as Solon Barocas and Helen Nissenbaum point out in Chapter 2, there is "overwhelming evidence that most of us neither read nor understand them."

The challenge presented by big data lies not just in the size of the data sets or the ubiquity with which data are collected but also, as Paul Ohm points out in Chapter 4, in the fact that big data "thrives on surprising correlations and produces inferences and predictions that defy human understanding." One can readily agree when he asks, "How can you provide notice about the unpredictable and unexplainable?"

In a May 2014 report (*4*), the U.S. President's Council of Advisors on Science and Technology described the framework of notice and consent as "unworkable as a useful foundation for policy." The report goes on to add that "only in some fantasy world do users actually read these notices and understand their implications before clicking to indicate their consent."

The authors of *Privacy, Big Data, and the Public Good* are generally unwilling to go so far. Despite pages of criticism, they advocate for tweaking the current system so that notice and choice are used more effectively. In Chapter 2, for example, Barocas and Nissenbaum propose that notice and consent be required in cases when the proposed use of the data departs from the context in which the data were collected. Alternatively, in Chapter 4, Ohm proposes that groups similar to institutional review boards might develop ethical norms that would determine when consent was necessary.

The book's critique of de-identification is equally compelling. Currently, de-identified data are exempt from most privacy laws. Yet throughout the book, the authors point to a number of well-publicized examples in which de-identified data sets have been released, only to have the data re-identified within days. In short, the more data are available, the harder it is to de-identify them effectively. Or, as Cynthia Dwork succinctly sums up the situation in Chapter 14: "'De-identified data' isn't."

On this subject, the authors offer an array of responses, ranging from Dwork's "differential privacy" approach, which uses statistical analyses to assess privacy risks presented by repeated queries of the same data set, to Ohm's recommendation that de-identified data should no longer be exempt from privacy laws.

*The reviewer is at the Indiana University Maurer School of Law, Bloomington, IN 47405, USA. E-mail: fred@fredhcate.org*

**Privacy, Big Data, and the Public Good**
**Frameworks for Engagement**
*Julia Lane, Victoria Stodden, Stefan Bender, and Helen Nissenbaum, Eds.*
Cambridge University Press, 2014.
342 pp.

There is a third aspect of the book that warrants special mention—a series of chapters addressing the uses of big data in the context of cities, urban planning, and research. These areas are not often discussed in the context of big data, and although they make the book feel a bit eclectic, they address new and interesting issues that clearly merit attention.

As society becomes more "datafied" (*5*)—a term coined to describe the digital quantification of our existence—our privacy is ever more at risk, especially if we continue to rely on the tools that we employ today to protect it. *Privacy, Big Data, and the Public Good* represents a useful and approachable introduction to these important issues.

## REFERENCES AND NOTES

1. U.S. Federal Trade Commission, *Privacy Online: A Report to Congress* (FTC, Washington, DC, 1998).
2. The White House, *Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation* (White House, Washington, DC, 2012).
3. *Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation)*, Unofficial Consolidated Draft Text, 22 October 2013.
4. President's Council of Advisors on Science and Technology, *Big Data and Privacy: A Technological Perspective* (PCAST, Washington, DC, 2014).
5. V. Mayer-Schönberger, K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin, Boston, 2013).

10.1126/science.1261092

Published by AAAS