Comments on Privacy Docket No. 180821780878001

Susan Landau
Fletcher School of Law and Diplomacy
and Department of Computer Science, School of Engineering
Tufts University

October 24, 2018

I appreicate the opportunity to submit comments on Developing the Administration's Approach to Consumer Privacy. I am Bridge Professor of Cyber Security and Policy in the Fletcher School of Law and Diplomacy and the School of Engineering, Department of Computer Science, Tufts University.[1] I have previously served as a Senior Staff Privacy Analyst at Google, a Distinguished Engineer at Sun Microsystems, and a faculty member at Worcester Polytechnic Institute, the University of Massachusetts Amherst, and Wesleyan University. I have also served on various advisory boards, and in particular, was a member of the Information Security and Privacy Advisory Board from 2002-2008. The comments I present here are my own and do not represent the views of any of the organizations with which I am affiliated. My comments draw heavily on my 2015 article "Contol use of data to protect privacy."[2]

[1] Affliation for identification purposes only.

[2] Susan Landau, "Control use of data to protect privacy, *Science*, 347 (Jan. 30, 2015), pp. 504-506, DOI: 10.1126/science.aaa4961.

1 Notice and Consent is no Longer a Useful Operating Tool

The approach taken by the Committee on Science and Law by the Association of the Bar of the City of New York[3] and later codified in the Fair information Practices of relying on Notice and Choice functioned well in a world of that time. Consumer data was collected only periodically rather than on a minute-by-minute—and sometimes second-by-second—basis. Data collection occurred in relatively large chunks, and the collection was typically done in a way that was clear to the data owner. And processing on this data was conducted by relatively few entities. We are no longer in this world.

The biggest change is so-called "big data," which the PCAST report *Big Data and Privacy: A Technological Perspective* report describes as "big in the quantity and variety of data that are available to be processed."[4]

An equally big change is the level of data being collected, so-called microdata. Whether it is collecting the order of swipes on a phone screen or location data when a consumer is using a mapping application[5], the data is collected at such a frequency and in such small, apparently innocuous pieces, that a user would be unwilling—indeed, essentially unable—to thoughtfully respond to a notice and consent request for each individual data collection.

With all this data arriving in digital form, computer analysis is a natural fit. The resulting big data analytics then provides strong predictive capabilities, which is of high value to both government and business. It is also a perfect storm.

The now classic example is that of Target's ability to determine, based on her purchases, that a young woman was pregnant; the company figured this out before her family did.[6] And as privacy experts well know, while any one individual microdata point is unlikely to provide particularly intrusive

[3]Alan F. Westin, *Privacy and Freedom,* New York: Atheneum, 1967, at ix.

[4]Presidents Council of Advisors on Science and Technology, *Big Data and Privacy: A Technological Perspective" Report to the President,* May 2014 https://bigdatawg.nist.gov/pdf/pcast_big_data_and_privacy_-_may_2014.pdf. [last viewed October 21, 2018, p. ix.]

[5]Google Maps collects location information as often as ten times a minute; smartray05, "Location History Sampling Info," December 11, 2012, accessed October 31, 2016, https://productforums.google.com/forum/#!msg/maps/ldKaXij4c_0/qsnMfunamM4J.

[6]Charles Duhigg, "How Companies Learn Your Secrets," *New York Times,* Feb. 16, 2012, https://www.nytimes.com/ 2012/02/19/magazine/shopping-habits.html.

personal information, the aggregate of such information may be so. It may demonstrate, for example, a phone user is increasingly agitated in using the device or that a driver stops, not occasionally, but daily, at a bar on her drive home from work.

There are multiple problems here: lots of data collection, microdata collection, data aggregation, in which seemingly innocuous, and perhaps de-identified, data is combined to create fine-grained portraits of individuals. As I wrote in 2015, "Notice simply doesn't make much sense in a situation where collection consists of lots and lots of small amounts of information."[7]

The immediate costs of the failure of notice and choice are to the individual whose data has been repurposed in ways that they did not anticipate and which may cause them harm. Whether it is the teenage girl whose family learns of her pregnancy before she was ready to tell them,[8] or the Facebook user who is targeted with a political ad determined through data inappropriately shared with an unexpected third party[9], the harm to the user is tangible and real. But it is society that also bears long-term costs from such breaches of trust. When private information is shared beyond boundaries that users expect, it is not just the users who were harmed who decide to hold their private information more closely; others do so as well.

The cost then accrues to society. As the PCAST report made eminently clear, the data collection and analytics are remarkably useful. Use of big data can guide urban planning, from where to move commuter bikes to respond to usage, to longer term planning of transportation options, zoning, etc. Use of microdata can guide a company towards developing a better user interface for a phone screen or provide information that enables a mapping application to suggest alternate routes when crowdsourced data shows a traffic bottleneck. The answer is not to prevent the collection of such microdata, but how to enable collection in a way that protects the user's privacy while enabling intended use of the information that is being provided.

[7]Landau *surpa* note 2, at 504.

[8]Duhigg *supra* note 6.

[9]See, for example, Carole Cadwalladr and Emma Graham-Harrison, "Revealed: 50 million profiles harvested for Cambridge Analytica in major data breach, *The Guardian,* March 17, 2018.

2 Protect Privacy through Controlling Use

In some sense, what underlies the Fair Information Practice Principles is the notion that individuals should be able to control how data they have shared about themselves is used. This is laid out specifically in the Purpose Specification principle, but is really embedded by the set of principles. The idea is that protecting privacy is done through controlling use of data. This principle is a pragmatic way of enforcing contextual privacy; the information a person supplies is dependent on context, and only the person is in a position to evaluate whether a different context satisfies their privacy expectations.[10] As the PCAST report put it, "The appropriate use of both the data and the analyses are highly contextual."[11]

The question is how to get from here—big data, microdata, data everywhere (data aggregation)—to protecting the privacy of users. A crucial component is on controlling use; there are a number of different approaches that do so.

One way is through the law. The first US example of controlling data use is, of course, the 1970 *Fair Credit Reporting Act*[12] (FCRA) which enables consumers to control dissemination of their credit information. FCRA has strict limits on the circumstances under which a person's credit information can be accessed (essentially for credit, employment, and in response to court orders).[13] The 2008 *Genome Information Nondiscrimination Act* (GINA) protects against discrimination in health insurance and employment based on genetic data.[14]

Another way is through legal action; a particularly striking case is that of the Havasupai tribe of northern Arizona, which successfully prevented the use of blood samples taken for one purpose—a diabetes study—from being used for another.[15]

---

[10]The theory of contextual privacy is due to Helen Nissenbaum. See, e.g., *Privacy in Context: Technology, Policy, and the Integrity of Social Life,* Stanford Law Books, 2009.

[11]PCAST *supra* note 4 at 47.

[12]15 U.S.C. §1681.

[13]Landau *supra* note 2 at 505

[14]GINA does not address discrimination against a person's relatives who may share the same genetic-induced disability. This is an enormous gap in the law, especially as an individual's genetic information may be revealed by *other people's* genetic information. This topic is, however, out of scope for this response.

[15]HAVASUPAI TRIBE OF the HAVASUPAI RESERVATION, a federally recognized Indian tribe, Plaintiff/Appellant, v. ARIZONA BOARD OF REGENTS and Therese

Members of the tribe were concerned about the increasing number of cases of diabetes. They talked with John Martin, an Arizona State University (ASU) anthropology professor who had worked with the tribe for decades and had the tribe's trust. In turn, Martin sought the help of ASU genetics professor Therese Markow. She agreed to do the study, but also expressed interest in expanding the genetic study to other diseases, including schizophrenia. Martin explained that the Havasuapi tribe would likely not be interested in an expansion of the study past diabetes.[16]

Between 1990 and 1992, members of the tribe provided blood samples and signed informed consent agreements enabling researchers to study whether a growing diabetes "epidemic" had genetic origins. When Martin learned in 2002 that the study had grown past the original intent of determining whether genetics explained the diabetes epidemic among tribal members to include other diseases, he informed the tribe. After attempting to negotiate with ASU, the tribe took the university and Markow to court. The Havasupai sought return of the DNA samples. The university settled out of court, with the DNA samples returned to the tribe.[17]

As Larry Lessig famously observed, "Code is law."[18] The architecture of Shibboleth, an identity-management system for sharing protected online resources, is an example of software that protects the identity of readers requesting an online resource from another institution.[19] If a member at an institution that participates in Shibboleth requests electronic resources from another institution—say a user at Tufts University seeks resources at Katholieke Universiteit Leuven (KU Leuven)—the user authenticates herself at Tufts. Tufts University knows the user's identity, but she is identified to KU Leuven not by her user name or her Tufts ID number, but rather by her right to access the resource (perhaps as a student in a particular course, a staff, student, or faculty member at Tufts, etc.). Only if the reader violates the rules under which the resource is borrowed is her identity revealed outside her home institution. This privacy-protective model is the result of

---

Ann Markow, Defendants/Appellees, Nos.1 CA-CV 07-0454, 1 CA-CV 07-0801. Decided: November 28, 2008.

[16]HAVASUPAI TRIBE OF the HAVASUPAI RESERVATION v. ARIZONS BOARD OF REGENTS and Therese Ann Markow *supra* note 15.

[17]Jennifer Couzin-Frankel, "DNA Returned to Tribe, Raising Questions About Consent," *Science,* Volume 328, Issue 558 (April 30, 2010), pp. 558.

[18]Lawrence Lessig, *Code and Other Laws of Cyberspace,* Basic Books, 1999.

[19]See Shibboleth Consortium, https://www.shibboleth.net/.

two factors: the Family Educational Rights and Privacy Act, which protects the privacy of student educational records and the importance that librarians place on reader privacy.

In setting up the rules governing Single Sign-On (SSO) systems that access federal sites, the federal government chose to use code—legal code[20]—to enable users to control access to their private activities while on federal sites. The tool was on online identity management systems.

Such systems are used across the Internet to access restricted resources, post comments, and conduct financial transactions. Development of Single Sign-On began nearly two decades ago, but adoption has been slow. In 2011, the federal government decided to jumpstart acceptance by creating the National Strategy for Trusted Identities in Cyberspace (NSTIC). The organization supplied funding for pilot programs and standards. But not only was NSTIC intended to give SSO capabilities a boost; the plan was to use SSO systems that worked in a privacy-protective manner. NSTIC requires that any private-sector identity providers used for accessing federal sites protect the privacy of user activities on those sites. That meant identity providers could not use tracking information from federal sites for anything but authentication, audit, or complying with the law.[21] Putting it another way, users could not be served ads, the identity providers could not share information about activities on the federal sites with a third party nor use the information to promote the identity provider's products. This is a high standard of privacy protection. Thus a signed-on user would have greater guarantees of privacy protections when visiting the National Cancer Institute website than the same user would have when visiting the American Cancer Society site.[22]

3 Recommendations

It is the case that protecting privacy through a broad stroke of Notice and Choice is easy from a process standpoint; require all data-collecting and data-using organizations to post such notices and receive user agreement.

[20]In this case, the code was regulations rather than laws.

[21]Georgia Tech Research Institute, GTRI NSTIC Trustmark Pilot, 2014, https://trustmark.gtri.gatech.edu/operational-pilot/trustmark-definitions/ficam-privacy-activity-tracking-requirements-for-csps-and-bae-responders/1.0/.

[22]Landau *supra* note 2 at 505.

But as many have observed, Notice and Choice no longer work effectively.y In recommending that, "[P]olicy focus primarily on whether specific uses of information about people affect privacy adversely," the PCAST report had the issue of privacy protection in an age of big data, microdata, and data aggregation exactly right.[23]

Unfortunately, despite some trenchant analysis of the "pressure points" (e.g., data collectors, data analyzers, users of the analyzed data[24]) within the big data ecosystem, there has been little follow up of the PCAST recommendations. I suspect the reason has to do with a confluence of events, including the rise of machine learning and the various serious cyberattacks (Ukraine power grid, U.S. election) in the interim. Attention has moved elsewhere, and the insights from this excellent insightful report appear to have dropped into an abyss. Thus my first recommendation is that the PCAST report, and its recommendations on controlling use, be carefully studied. These recommendations are useful.

Use cases are not an easy sell; unlike Notice and Choice, use cases depend on the particularities of the situation. Reliance on use cases for protecting privacy does not provide a simple broad sweep solution in the same way that Notice and Choice appeared to do for many years. That is not a reason not to rely on use cases—given the complexity of the way we use data these days, we must do so—but rather an observation that reliance on use cases for protecting privacy will require significant detailed study in order to develop methodology and put its use to good effect.

On the policy-development side, use cases will depend on detailed study of cases. It appears, for example, that people suffering from cancer are far more willing to share information about treatment and outcomes with researchers than is the case with patients suffering from other diseases.[25] If this is correct, then how do we design protocols—computer protocols, not treatment protocols—in the two sets of situations for participants who are generally willing to share information on treatment and outcomes, and for those who are generally unwilling to share such data?

Thus I reccomend several steps. We need to understand how targeted policy solutions can provide privacy and we need to develop techncial solutions, such as is used, for example, by Shibboleth, that can do the same. It is

[23]PCAST *supra* note 4 at v.

[24]PCAST *supra* note 2 at 47-49.

[25]Conversation with Eric Lander, approrimately 2007.

no surprise then that I suggest more study is needed. I have several specific recommendations:

1. I recommend that the National Academies of Science, Engineering and Medicine be asked to further the work of the PCAST report and conduct a study into how to employ use cases to better protect privacy. This study should examine in depth both policy (e.g., as was done by the U.S. government for serving as an identity provider to government websites) and technical solutions, as well as their combination.

   We have two domains, personal financial records and medical records, in which we have some experience in providing legal and policy protections for protecting personal data. Thus it is likely to be valuable for such a study to examine these particular domain areas in detail.

2. I recommend that the National Science Foundation put out a solicitation for work in developing privacy protections in specific use cases much in the style that was done by Shibboleth, a targeted solution for a specific problem.

3. There is also the possibility of developing use-control techngologies that work more broadly than a Shibboleth type solution. For example, MIT researchers Oshani Seneviratne and Lalana Kagal created a protocol, httpa, that was designed to track information usage across websites.[26] The protocol does not provide enforcement, but does enable determining when data usage policies have not been followed. This solution does not appear to be easy to scale—but more study would be useful. Even better would be study to develop other tools for enforcing usage possibilities.

Four years late—but better late than never—we need to follow up on the privacy solutions proposed in the PCAST study by seriously looking at use cases. If we are to protect consumer privacy in the age of big data, microdata collection, and data aggregation, we have no other choice.

[26]Oshani Seneviratne, Lalana Kagal, *IEEE International Symposium on Policies for Distributed Systems and Networks,* 2011, pp. 141144.