Privacy and Big Data

Brian M. Gaff, Heather Egan Sussman, and Jennifer Geetter,
McDermott Will & Emery, LLP

Big data's explosive growth has prompted the US government to release new reports that address the issues—particularly related to privacy—resulting from this growth.

The White House just released two reports addressing the public policy implications of big data's proliferation. The first report is by the President's Council of Advisors on Science and Technology and is entitled, "Big Data and Privacy: A Technological Perspective" (PCAST report; www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf). The second, which relies in part on the PCAST report, is by the Executive Office of the President and is entitled, "Big Data: Seizing Opportunities, Preserving Values" (White House report; www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf). Both reports make the case that instead of slowing the accumulation of data or placing barriers on its use, we should focus on policy initiatives and legal frameworks that foster innovation, promote the exchange of information, and support public policy goals, while at the same time limiting harm to individuals and society.

For an expanded discussion on these issues, listen to the podcast that accompanies this column at www.computer.org/portal/web/computingnow/computing-and-the-law.

BACKGROUND ON THE REPORTS

The PCAST and White House reports were released simultaneously on 1 May 2014 following a 90-day study on big data. The reports contain an analysis of whether the US can forge intentional standards on how to manage this data and how we can continue to promote the free flow of information in ways that meet both privacy and security requirements.

The PCAST report

The PCAST report provides technical guidance on big data information technology and the privacy challenges presented by collecting, storing, and leveraging big data assets. It characterizes data that's "born digital" (created for use specifically for a computer or digital processing system, such as email), and "born analog" (created from elements of the physical world captured in digital form, such as faces or voices through sensors like cameras or recording devices).

For data that's born digital, the PCAST report identifies two potential privacy concerns: overcollection and data fusion. *Overcollection* happens when engineering or device design causes the collection of data unrelated to the stated purpose. An example of overcollection is a social networking app that clandestinely collects a user's contact list and then spams those contacts with advertisements. *Data fusion* occurs when data from different sources are brought together and new phenomena emerge. Examples of this include data mining and pattern recognition. A privacy challenge in the case of data fusion is how to apply traditional frameworks such as "notice and consent" when data fusion happens in unexpected ways

0018-9162/14/$31.00 © 2014 IEEE

and produces unexpected results. The PCAST report doesn't attempt to answer how to address this privacy challenge but notes that it must be a priority going forward.

The PCAST report has four key conclusions:

- Encryption isn't a perfect solution for securing big data, but it could be a valuable component in a comprehensive privacy solution.
- Third parties would create various privacy profiles for consumers who would then select a profile such that data holders would be required to differentiate the way they use data based on each consumer's selection.
- Anonymization and de-identification have limited relevance because data points linked to one another tend to take on other identifiable attributes.
- Deletion and nonretention policies aren't effective means of protecting individual privacy.

The PCAST report also makes five recommendations:

- The focus should be on the actual uses of big data and not so much on its collection and analysis.
- To avoid obsolescence, policies and regulations should be stated in terms of intended outcomes and not embed particular technological solutions.
- The US should strengthen its research in privacy-related technologies.
- There should be more education and training opportunities concerning privacy protection.
- The US should take the lead by adopting policies that stimulate the use of practical privacy-protecting technologies that exist today.

The findings and recommendations described in the PCAST report underpin the White House report's technology discussion and conclusions.

The White House report

The White House report makes clear that the value of big data analytics has been established and that going forward, the focus should be on how to allow big data analytics to proceed unimpeded while still protecting privacy and other important values. The report doesn't define big data; instead, it describes it using the well-known "three V's" model that connects volume, variety, and velocity. In other words, big data is information that's so large in volume, so diverse in variety, or moving with such velocity that traditional modes of data capture and analysis are insufficient.

The White House report raises critically important questions about whether sufficient mechanisms are in place to protect our privacy and other important values.

While acknowledging the seemingly unlimited future uses of accumulated data, the White House report raises critically important questions about whether sufficient mechanisms are in place to protect our privacy and other important values.

To start addressing these issues, the report makes six policy recommendations to promote the responsible and accountable use and disclosure of big data:

- Advance the Consumer Privacy Bill of Rights to promote the responsible use of big data.
- Pass national data-breach legislation that includes reasonable time periods to notify individuals affected, minimizes interference with law enforcement investigations, and prioritizes notification for large harmful incidents over smaller ones.
- Expand privacy protections to non-US citizens.
- Ensure data collected on students in school is used for educational purposes.
- Expand technical expertise to stop discrimination by using new ways of analyzing big data to detect and investigate it.
- Amend the US Electronic Communications Privacy Act to ensure that electronic data—such as metadata—is afforded protections consistent with those afforded to physical items.

WHAT COMPANIES SHOULD DO NOW

Both the PCAST and White House reports endorse and emphasize the Consumer Privacy Bill of Rights, which embodies certain principles including transparency, respect for context, security, access and accuracy, focused collection, and accountability. Consequently, companies that participate in the big data ecosystem—those that facilitate, collect, process, analyze, use, or benefit from big data—should consider appointing a team that's responsible for three key areas: understanding the Consumer Privacy Bill of Rights and its underlying principles; analyzing how existing internal systems, policies, procedures, and practices align with—or might need to be adjusted to reflect—these principles; and evaluating how new products and business models incorporate and reflect these principles, so as to stay ahead of where the potential regulations are headed.

In addition, companies should designate a team to monitor any further activity on these two reports, assessing the business impact of any associated actions by industry, the White House, related agencies, or lawmakers. Such a team would be able to advise on whether lobbying efforts are appropriate to make the company's voice heard during the evolving discussions on these issues.

Companies should be creative as well. Given the reports' caution about the potential limits on the "notice and consent" framework as a panacea for privacy controls in a big data environment, and in light of ongoing US Federal Trade Commission consideration of how to protect consumer-generated digital information, software developers, engineers, app creators, and other innovators should think about tools that can be integrated into their products to provide greater control and flexibility downstream as new norms and expectations regarding privacy, transparency, and accountability come into focus.

Both the PCAST and White House reports acknowledge that there's still much more work to be done, many debates to be had, and more stakeholder input to consider in the weeks and months ahead. There's no question that with these two reports, the White House has firmly planted a stake in the ground, declaring big data as the "new normal" and making clear that businesses, public policy, and legal frameworks must adjust. The adjustments will likely be comprehensive, so get advice from your lawyer to ensure that you comply with what will probably be a rapid change in regulations. C

Brian M. Gaff *is a senior member of IEEE and a partner at the McDermott Will & Emery, LLP law firm. Contact him at bgaff@mwe.com.*

Heather Egan Sussman *is a partner at the McDermott Will & Emery, LLP law firm. Contact her at hsussman@mwe.com.*

Jennifer Geetter *is a partner at the McDermott Will & Emery, LLP law firm. Contact her at jgeetter@mwe.com.*