White House-MIT Big Data Privacy Workshop[1]
A Personal View

By
Dr. Michael L. Brodie
Computer Science and Artificial Intelligence Laboratory, MIT
March 24, 2014

Actual and potential societal benefits of information in our increasingly digital world are leading to significant shifts in values (e.g., security versus freedom) in behaviours, and ultimately in polices and laws. More than ever before, technology is far ahead of the law. While the scale of Big Data offers massive potential, that very scale can render infeasible conventional, top-down[1] solutions, including security technologies and methods. We are in the midst of two significant shifts – the shift to Big Data requiring new computational solutions, and the more profound shift in societal benefits, risks, and values.

The White House clearly enunciated the benefits and threats of Big Data. President Obama asserted for this study the societal objective of *Keeping people safe in a world of Terrorism while maintaining their individual privacy* in conjunction with the technical objectives that *the Internet and Big Data must be open and interoperable*. This was reinforced by the Honorable Penny Pritzker, Secretary of the US Department of Commerce, who recognized the potential of Big Data to enhance innovation, productivity, and economic growth based on the *free flow of information* and on *mutual trust amongst citizens, corporations, and governments*. These objectives underlie the White House study to determine: *Whether our existing privacy framework can accommodate these changes, or if there are new avenues for policy that we need to consider*. John Podesta, Obama's Counselor and study lead asked: *Have we fully considered the myriad ways in which this data revolution might create social value*, he added, *and have we fully contemplated the risks that it might pose to our conceptions of individual privacy, personal freedom and government responsibility of data*? *Does our legal privacy framework support and balance safety and freedom*?

The White House-MIT Big Data Privacy Workshop: *Advancing the State of the Art in Technology and Practice*[1], (the first of three such workshops[2]) responded to the technical questions. The emergence of Big Data and its applications pose significant technical challenges not only in implementation and operation but also in managing the risks for which there are emergent and as yet incomplete solutions such as Dwork's differential privacy, and emerging top-down solutions like Zeldovich's encrypting databases (CryptDB) and Vadhan's computing on encrypted data. Due to the massive scale of Big Data, previously top-down solutions for security, e.g., anticipating and preventing security breaches, will simply not scale to Big Data. They must be augmented with new approaches including bottom-up solutions such as Stonebraker's logging to detect and stem previously unanticipated security breeches and Weitzner's accountable systems. *"Big data" has rendered obsolete the current approach to protecting privacy and civil liberties[3]*. Hence, Big Data requires a shift from a focus on top-down methods of controlling data generation and collection to data usage. Not only do top-down methods not scale, *Tightly restricting data collection and retention could rob society of a hugely valuable resource[3]*. Adequate let alone complete solutions will take years to develop.

Yet technical solutions must be designed to meet requirements that are not yet fully known, such as the societal value of privacy in an increasingly digital world that has the potential of recording our every action and thought[2]; and establishing and maintaining trust amongst citizens, corporations, and governments both at home and abroad, since an open Internet knows no sovereign boundaries. Perhaps the most compelling idea that emerged from the workshop was stated by Guttag who observed that the

---

[1] Top-down versus bottom-up corresponds loosely to theory-driven versus data-driven.

[2] The Wall Street Journal estimates that the average Internet user is tracked in excess of 2,000 times per day. Madden indicated that 5 billion cellphones track every move of their users.

importance of safe, secure, efficient solutions should not stand in the way of urgent challenges that could leverage Big Data solutions. Ask the families of critically ill patients how much security they would risk to extend the length and quality of the lives of their loved ones by exploiting the potential of Big Data.

Many other challenging questions were posed without answers, including, *What is the economical value of data*? Is it an asset like money? How does such valuation change how information is managed?

In summary there was broad agreement on the societal and technical objectives of the White House study and on encouraging technical and research directions. While legislation has always lagged technology advances; the gap, evidenced by recent NSA surveillance activities, is at an extreme, at least in terms of public perception. What was least understood at the workshop, and more broadly, was the balancing of security and freedom in values and then in legislation and the impact on our democracy and the ways in which we want to live and work.

Big dominated the workshop: Big Data, Big Trust/Liability, Big Rules, Big Compliance, and ultimately *Big Shifts* in values, methods, and regulation to obtain Big Benefits and keep up with Big technology advances like Big Data.

At the South By Southwest Conference a week after the White House-MIT Workshop, Edward Snowden addressed in a virtual conversation [5] safety and freedom in the face of his disclosures of NSA surveillance, the very actions that, in part, prompted the White House Study. Remarkably, the White House and Podesta [1] and Snowden [5] were in almost complete agreement on the objectives and challenges. In the workshop both the White House and Snowden addressed the issues objectively and were articulate and credible. However, the same could not be said a week later in Snowden's TED Virtual Conversation [6] in which Snowden's position was more assertive on democratic principles such as liberty, privacy, and transparency for an open Internet and on the same technical issues raised earlier, again articulate and credible, this time supported by the inventor of the World Wide Web, Sir Tim Berners-Lee. Unfortunately, in his Response to Snowden's Virtual Conversation [7], Richard Ledgett, Deputy director NSA, did not address the technical issues raised by the White House and Snowden preferring to comment on political issues and Snowden's motivations and culpability, seemingly not aligned with the attitude expressed by the White House Study, the White House-MIT workshop, nor the SXSW and TED audiences.

These public actions of President Obama, the White House, and relevant government agencies, striving for an open and free Internet while balancing safety with freedom is in stark contrast to corresponding government actions around the world, including Turkey, Brasil, China, and even Switzerland. As Big Data opens the doors to a new world of computing, we should embrace the opportunities and understand the risks rather than close the door. If we don't, others certainly will.

References

[1] The White House-MIT Big Data Privacy Workshop: Advancing the State of the Art in Technology and Practice

The White House Office of Science and Technology Policy (OSTP) and MIT co-hosted a public workshop entitled "Big Data Privacy: Advancing the State of the Art in Technology and Practice" on March 3, 2014. The event was part of a series of workshops on big data and privacy organized by the **MIT Big Data Initiative at CSAIL** and the **MIT Information Policy Project**. The workshop was also the first in a series of events being held across the country in response to President Obama's **call for a review of privacy issues** in the context of increased digital information and the computing power to process it.

[2] White House Workshop Series

As part of this effort, OSTP will be co-hosting at least two additional events—one with the Data & Society Research Institute and New York University, and one with the School of Information and the Berkeley Center for Law and Technology at the University of California, Berkeley. In the coming weeks, we will be announcing additional opportunities for the public to inform this important work. Check back here for more information and updates on our progress.

[3] Craig Mundie**, Privacy Pragmatism:** Focus on Data Use, Not Data Collection, Foreign Affairs, March/April 2014.

[4] Keynote speaker: John Podesta, White House Counselor, White House-MIT Big Data Privacy Workshop: Advancing the State of the Art in Technology and Practice, March 4, 2014, MIT, Cambridge, MA *http://web.mit.edu/bigdata-priv/agenda.html*

[5] Edward Snowden, "Virtual Conversation with Edward Snowden" SXSW, Austin, TX Monday, March 10, 2014, https://www.youtube.com/watch?v=UIhS9aB-qgU and http://sxsw.com/interactive/news/2014/snowden-sxsw-2014-watch-video-historic-march-10-session-here

[6] Edward Snowden, Here's how we take back the Internet, TED, Vancouver, Canada, March 19, 2014
https://www.ted.com/talks/edward_snowden_here_s_how_we_take_back_the_internet

[7] Richard Ledgett, Deputy director, NSA; **The NSA responds to Edward Snowden's TED Talk,** TED, Vancouver, Canada, March 20, 2014, http://www.ted.com/talks/richard_ledgett_the_nsa_responds_to_edward_snowden_s_ted_talk