



Where the White House "Big Data" Report Falls Short

BY [Jessica McKenzie](#) | Tuesday, May 6 2014

The White House released its [report on big data](#) Friday to general approval from civil rights advocates for its acknowledgement of the dangers of discrimination through new ways of manipulating, combining and analyzing personal data. However, a number of concerns remain: that the report was too starry-eyed about big data; that the report gave preference to industry stakeholders rather than citizen consumers; and that its policy recommendations were not forceful enough.

As the Intercept's Marcy Wheeler characterized the report in [a post](#) last Friday, it's really "rather breathless."

The report, authored primarily by senior White House advisor John Podesta, opens with a generalization about the boon data has been to society "since...ancient times." Less savory uses of data collection—like the use of census data to identify Japanese-Americans during World War II before forcing them into internment camps—are buried in the middle pages.

And, as techPresident's Micah Sifry pointed out in [First Post last Friday](#), it skates over a politically sticky point: the use of big data "to better understand and potentially manipulate voters."

Discrimination—Front and Center

Chris Calabrese, legislative council for the American Civil Liberties Union, told techPresident that the fact the report even acknowledged the potential for discrimination in big data is huge. A footnote on the potential for digital redlining cites the "[Civil Rights Principles for the Era of Big Data](#)," five pillars of thought that the ACLU and 13 other organizations suggest should guide policy making.

When Calabrese spoke with techPresident about the Principles in March, he said "The Chicago heat map should make every American very nervous."

He was referring to the list of 400 individuals the Chicago police department determined through big data analysis to be most likely to be involved in a violent crime.

This example of predictive policing is mentioned in the White House report as being controversial. Although the report says that these practices merit review for their potential to

infringe on constitutional rights, it is also optimistic about their efficacy in preventing violent crime, leaving the reader a bit in the dark as to the authors' stance.

However, Calabrese told techPresident that just the fact that the report mentions the Chicago example is a really big deal.

Seeta Peña Gangadharan, a civil rights advocate and Senior Research Fellow at the New America Foundation's Open Technology Institute, shares Calabrese's optimism.

“We can no longer exclude these issues [of discrimination] from the conversation about big data,” she told techPresident.

Next Steps

The reports says that “additional research in measuring adverse outcomes due to the use of scores or algorithms [that results in [differential pricing](#) or targeted advertisements] is needed to understand the impacts these tools are having and will have in both the private and public sector as their use grows.”

Yet one of Gangadharan's lingering questions was what it would actually take for the government to perform a “nitty gritty disparate impact analysis.” Although it was a nice suggestion, she really wanted to know how it was going to get done.

Edward W. Felten, a computer scientist at Princeton and former chief technologist of the Federal Trade Commission, echoed the assertion that government and industry need to use data analysis to address digital discrimination.

“There is a role for government to hold companies accountable and establish incentives,” Felten [told the New York Times](#). “There needs to be enough incentive for companies to do the hard work.”

Gangadharan, however, still wonders what those incentives are or should be.

One thing Gangadharan did notice is that the report seemed invested first in innovation and economic growth, followed by privacy and discrimination considerations.

Joseph Turow, author of [The Daily You: How the New Advertising Industry is Defining Your Identity and Your Worth](#), observed something similar. “It nods to all the right stakeholders except the public,” he told techPresident.

“The key thing is that the report hits a lot of important buttons but I'm not convinced that it is hard-nosed enough about its conclusions,” Turow said. “[We need] true citizens' discussions about how the Internet works and the issue of privacy and how to reach out to people in government...[but] this report will have very little meaning to most people unless we figure out how to do this.”

“Fundamentally Intractable Things”

“The report too easily slides over things,” explained Turow. “It makes it seem that these things are not intractable...[when] there are some fundamentally intractable things.”

Turow points to one particular contradiction in the report, in which the Podesta wonders: “Should there be an agreed-upon taxonomy that distinguishes information that you do not collect or use under any circumstances, information that you can collect or use without obtaining consent, and information that you collect and use only with consent?”

“This contradicts the entire idea of big data,” Turow said. “Inherent in the problem of big data [is the idea that] the smallest benign bits of data can be aggregated to yield profound conclusions. Even if we agree upon what is acceptable [to collect] those things can be aggregated in ways that ultimately come up with unacceptable conclusions.”

The idea of intractability appeared in another form in Wheeler's blog post:

Perhaps the most frustrating part of the report...is the silence about the things we don't use Big Data for enough, notably solving the financial crisis and regulating banksters (including things like tax havens, inequality, and shadow banking), or really doing something about climate change.

Big Data, as it appears in the report (as presented by a bunch of boosters) is not something we're going to throw at our most intractable problems.

Distraction or Gesture?

Some experts have asserted that this report is a distraction from the National Security Agency's programs.

"It's an important issue to study," Cato Institute Research Fellow Julian Sanchez [told the Washington Post in February](#), "but given the odd timing, I think it read to many folks as—take your pick—a distraction, a bone to privacy advocates, or a shot across the bow of tech companies that have been pushing back on NSA."

(Incidentally, Edward Snowden is only mentioned in the report once, in a paragraph about the need for ongoing evaluations of government employees with security clearances. He and fellow whistleblower Chelsea Manning are lumped together with the Fort Hood and Washington Navy Yard shooters in a stunning equivocation over “troubling breaches” and “acts of violence.”)

However, the civil rights and privacy advocates I spoke with were all more concerned that overall, the report was a nice gesture that might not see political follow-through. Both Calabrese and Gangadharan asserted that the hard work was still to come.

“We need to do more auditing, we need more information about how [big data] is being used and new tools to create accountability,” Calabrese said. “We need to begin to take some of the burden off of consumers and we need to make clear that some practices are not ok.”

The Obama Administration released a set of suggested guidelines called the [Consumer Privacy Bill of Rights in 2012](#), and yet those recommendations have yet to be enshrined into law by Congress.

“Eventually,” Calabrese told techPresident, “reports are not going to be enough.”