



October 31, 2014

Submitted electronically via <https://ftcpublic.commentworks.com/ftc/bigdataworkshop>

Federal Trade Commission, Office of the Secretary
Room H-113 (Annex X)
600 Pennsylvania Avenue, NW
Washington, DC 20580

Re: “Big Data: A Tool for Inclusion or Exclusion” Post-Workshop Comments

The Internet Association¹ appreciates the opportunity to provide further comments regarding the Federal Trade Commission’s September 15 “big data” public workshop, “Big Data: A Tool for Inclusion or Exclusion?” Prior to this workshop, our association provided comments to the Commission in which we discussed the significant benefits of leveraging data analytics not only within the Internet industry but also across multiple sectors. Additionally, we expressed support for a conversation shift toward a responsible use framework of data and away from data collection through a self-regulatory framework.

In light of the discussions that occurred during the workshop, we submit this comment in order to reaffirm our position on “big data”: that its benefits and costs to society be weighed carefully so as to ensure that its potential is maximized while protecting consumers.

The Internet Association members are dedicated to protecting consumers online, and understand very clearly the importance of maintaining users’ trust and their own reputations as trusted platforms. Internet companies have long leveraged data analytics to provide useful

¹ The Internet Association represents the world’s leading Internet companies including: Airbnb, Amazon, AOL, Auction.com, eBay, Etsy, Expedia, Facebook, Gilt, Google, Groupon, IAC, LinkedIn, Lyft, Monster Worldwide, Netflix, Practice Fusion, Rackspace, reddit, Salesforce.com, Sidecar, SurveyMonkey, TripAdvisor, Twitter, Uber Technologies, Inc., Yelp, Yahoo!, and Zynga.



products and services directly benefiting not only users, but also society and the economy more broadly. A recent study found that in 2013, “big data” added significant value to the economy; more specifically, “big data” product and service sales exceeded \$18 billion, and are expected to grow to \$50 billion by 2017.² To support these economic benefits, companies are using data to address important employment issues. For instance, in an increasingly challenging environment for new college graduates entering the workforce, LinkedIn recently debuted new tools that use data-driven insights to help students better navigate this environment.³ These tools allow users to benefit from career paths of LinkedIn members and other social features to receive guidance from trusted advisors and professionals.⁴ Furthermore, Google recently relied on “hard statistics” to determine effective methods to encourage its female engineers to apply for promotions at higher rates.⁵

From a social good viewpoint, data’s important societal impact is currently being demonstrated through data scientists’ efforts to analyze large data sets, which will ultimately help prevent the Ebola virus from spreading.⁶ At the time of writing, the U.S. Centers for

² U.S. CHAMBER OF COMMERCE FOUNDATION, *THE FUTURE OF DATA-DRIVEN INNOVATION* (2014), available at <http://www.uschamberfoundation.org/sites/default/files/The%20Future%20of%20Data-Driven%20Innovation.pdf>.

³ Christina Allen, *Social +Data = Better Decisions for Students*, Oct. 1, 2014, LINKEDIN OFFICIAL BLOG, <http://blog.linkedin.com/2014/10/01/social-data-better-decisions-for-students/>.

⁴ *Id.*

⁵ Cecilia Kang, *Google data-mines its approach to promoting women*, WASHINGTON POST, Apr. 2, 2014, <http://www.washingtonpost.com/blogs/the-switch/wp/2014/04/02/google-data-mines-its-women-problem/> (explaining that Laszlo Bock, Google’s senior vice president of People Operations, shared the results of two studies finding that (1) “girls typically raise their hands less often than boys when answering math problems and (2) women usually offer ideas in business meetings less than men with its female engineers in an email reminder about applying for promotions”).

⁶ Matthew Wall, *Ebola: Can big data analytics help contain its spread?* BBC NEWS, Oct. 14, 2014, <http://www.bbc.com/news/business-29617831>.



Disease Control and Prevention (CDC) is aggregating mobile phone data from mobile operators to trace specific regions where people are contacting Ebola helplines.⁷ This data is being used to operationalize the CDC’s response on the ground so as to effectively deploy resources to the areas most in need.⁸ At the state level, cities are also relying on data to better understand important issues such as crime. For instance, the city of New Orleans established its first-ever innovation delivery team (also known as the “i-Team”) to analyze its crime and murder data from the 1960s to date.⁹ To reduce its crime rate and to support those most at risk of becoming both victims and perpetrators, the city used its findings to develop its “NOLA for Life” strategic plan consisting of five pillars: stop the shooting, invest in prevention, promote job growth, rebuild neighborhoods, and strengthen city law enforcement.¹⁰

These social good initiatives often rely on corporate data.¹¹ Recently, MIT’s Media Lab announced the launch of its Laboratory for Social Machines (“Lab”) backed by Twitter.¹² The social media platform will give the Lab access to both its archive of public tweets and its real-

⁷ *Id.*

⁸ *Id.*

⁹ Jason Shueh, *New Orleans Cuts Murder Rate Using Data Analytics*, Oct. 22, 2014, SOLUTIONS FOR STATE AND LOCAL GOVERNMENT – GOVERNMENT TECHNOLOGY, <http://www.govtech.com/data/New-Orleans-Cuts-Murder-Rate-Using-Data-Analytics.html>.

¹⁰ *Id.*

¹¹ For example, MIT’s Computer Science and Artificial Intelligence Laboratory executed a deep learning project in which Google Street View images were fed into an algorithm to accurately make inferences such as the nearest McDonald’s in people’s vicinity and the crime rate of an area. Fourth year PhD student Aditya Khosla, who worked on the project, anticipates that these initial findings indicate greater positive implications for society. For instance, the research team predicts that research applications could range from determining the best locations for schools or hospitals during town planning to drawing conclusions about parts of the world that do not regularly publish crime or other statistics. Khosla notes that the societal benefits that could result from this research rely on more data. *See generally* Luke Dormehl, *This Algorithm Predicts A Neighborhood’s Crime Rate Using Google Street View*, FAST COMPANY, Oct. 16, 2014, <http://www.fastcolabs.com/3036677/this-algorithm-knows-your-neighborhood-better-than-you-do>.



time, public tweets with the intended goal of creating new platforms for users and institutions to identify and address complicated social issues.¹³ Additionally, cities like Portland, Oregon are relying on data collected via mobile applications to improve safety for its residents.¹⁴ More specifically, the Oregon Department of Transportation (ODOT) licensed a dataset consisting of about 17,000 riders and 400,000 bike trips around Portland from Strava - a GPS-powered application that tracks cyclists and runners' workout progress – to make the city safer for riders and cyclists.¹⁵ While the disaggregated and anonymized data faces challenges with sample bias, ODOT believes that this presents an opportunity for learning about ways to improve cities for their residents.¹⁶

To encourage these critically important social benefits of “big data,” we continue to urge policymakers and regulators to support and maintain the United States' existing privacy regime, which provides the ability for companies to freely innovate while ensuring that consumers are protected online. During the workshop, Chairwoman Ramirez outlined post-workshop actions and alluded to conducting a review of existing law - both privacy and non-privacy related - to determine any potential gaps.¹⁷ Additionally, the White House's “big data” report, “Big Data: Seizing Opportunities, Preserving Values (White House Big Data Report)” also called for a review of the future of privacy protections, including concepts such as notice and consent.¹⁸

¹² MIT Media Lab, *MIT Launches Laboratory for Social Machines with Major Twitter Investment*, MIT NEWS, Oct. 1, 2014, <http://newsoffice.mit.edu/2014/twitter-funds-mit-media-lab-program-1001>.

¹³ *Id.*

¹⁴ Mona Lalwani, *Portland enlists big data to make biking safer*, THE VERGE, Oct. 8, 2014, <http://www.theverge.com/2014/10/8/6938041/portland-enlists-big-data-to-make-biking-safer>.

¹⁵ *Id.*

¹⁶ *Id.*

¹⁷ Transcript of Federal Trade Commission – Big Data: A Tool for Inclusion or Exclusion (Sept. 16, 2014), http://www.ftc.gov/system/files/documents/public_events/313371/bigdata-transcript-9_15_14.pdf at 11 [hereinafter *Transcript*].

¹⁸ EXECUTIVE OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 61 (2014), *available at* http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.



To avoid compromising future data innovation, we support this comprehensive review in order to better inform the debate and better identify potential gaps before preemptive action is taken.

This methodical approach is a necessary precursor to considering the potential harms associated with “big data.” At the outset of this workshop, the Commission sought to “shed light on the full scope of big data practices ... and examine the *potentially* positive and negative effects of big data...”¹⁹ It is clear from the White House Big Data Report and from some workshop panelists’ remarks, that big data benefits are a reality of the current marketplace—not merely prospective. In contrast, the negative effects of “big data” are still largely speculative. Any gaps analysis should focus on actual, concrete harms that are occurring, and should also delineate privacy harms and discrimination harms separately, as their remedies are found in distinct bodies of law.

One impetus for the workshop is the Commission’s concern regarding potential “discrimination by algorithm.” One panelist, Solon Barocas, Ph.D, presented on the mechanics of machine learning and posited that data mining is a form of statistical discrimination.²⁰ We urge the Commission to shift its focus away from technology itself to the uses and misuses of information—regardless of the technology employed. Technology in and of itself is neutral; it is the values with which it is imbued by human actors that can lead to discrimination. This topic should be addressed as part of a larger societal discussion around evolving modes of discrimination. By limiting the focus to discrimination by algorithm, we risk failing to account for the many new and changing forms of discrimination that occur outside the big data context. Using information to make detrimental determinations about individuals’ credit, health, and

¹⁹ Federal Trade Commission, Big Data: Tool for Inclusion or Exclusion?, <http://www.ftc.gov/news-events/events-calendar/2014/09/big-data-tool-inclusion-or-exclusion> (last visited Oct. 15, 2014).

²⁰ *Transcript* at 19.



employment should be subject to the same level of policy scrutiny, regardless of whether this is accomplished through technological or manual methods. As this discussion continues, we encourage the Commission to work in conjunction with the federal government’s civil rights and consumer protection agencies as well as with academics, subject matter experts, and other qualified parties to explore the landscape of existing anti-discrimination laws and explore the potential for new, harmful uses to determine how best to address them.

Companies are using data analytics to decrease discriminatory practices. For example, big data firms are relying on data to eliminate biases that may occur when hiring employees by adjusting their algorithmic formulas to discard certain factors that could unintentionally discriminate against certain groups.²¹ In addition, to promote a more inclusive society, the public and private sectors support data uses that rely on automated processes to reduce human biases in decision-making, encourage people to treat others based on their individual attributes rather than stereotypes, and change controversial policies such as New York’s stop-and-frisk policy.²² Therefore, we urge the Commission to recognize discriminatory practices as technology-neutral while supporting a conversation shift towards responsible uses of data.

Finally, in addition to transitioning the debate towards responsible uses of data through a self-regulatory approach, we urge the Commission to explore additional factors such as:

- Transparency: In explaining the intersection of “big data” and consumer protections during the Commission’s workshop, Commissioner Brill stated that, “Consumer trust is critical here

²¹ Dustin Volz, *How Big Data Can Be Used to Fight Unfair Hiring Practices*, National Journal, Sept. 26, 2014, <http://www.nextgov.com/big-data/2014/09/how-big-data-can-be-used-fight-unfair-hiring-practices/95195/>.

²² Daniel Castro, *Big Data is a powerful weapon in the fight for equality*, THE HILL, Oct. 23, 2014, <http://thehill.com/blogs/pundits-blog/technology/221583-big-data-is-a-powerful-weapon-in-the-fight-for-equality>.



and transparency and accountability are key to building it.”²³ The Internet Association’s member companies support Commissioner Brill’s calls for transparency and accountability for consumers. Our industry believes in the power of online tools that provide consumers with the ability to learn more about how their data is used. Internet platforms have been working hard to develop tools that provide users access to their information, and will continue to refine these tools and build new mechanisms over the coming years. For instance, Facebook recently introduced a new transparency and control tool called Ad Preferences. This tool allows people to get information about why they are seeing any individual ad on Facebook and also gives people the ability to see the interests Facebook has associated with them. Ad Preferences gives people direct control and allows them to view their accounts at a granular level to either remove interests that they do not prefer, or add new interests if they want to curate their ad experiences. Additionally, Google’s Dashboard offers users the ability to view information regarding how they use Google’s product, as well manage their settings to adjust how users engage with each product (e.g., clear recent search history), and monitor their account for suspicious activity such as unusual geographic locations or irregular spikes in emails sent from their account.²⁴ Further, Yahoo, has long offered users its Ad Interest Manager (AIM) tool, which gives users transparency into the standard interest categories the search company uses to help it choose the kinds of advertisements and content users see. Through AIM, users can edit or deselect categories, or opt out of interest-based ads, content and other features altogether. Based on robust access tools like the ones described above, Internet users are empowered to determine how they want to interact with services and content on the platforms of their choice. In other words, our member companies already provide end users with tools that facilitate transparency and accountability online.

²³ *Transcript* at 141-42.

²⁴ Google, *View account history in Google Dashboard*,

<https://support.google.com/accounts/answer/162744?hl=en> (last visited Oct. 29, 2014.).

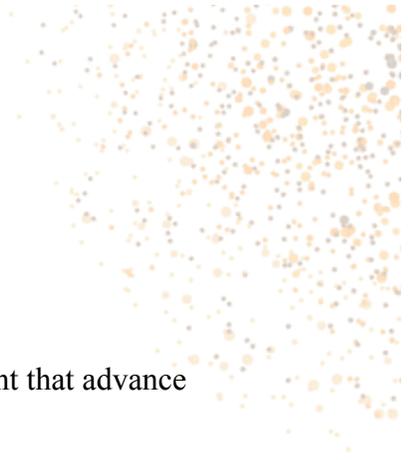


- Cost-benefit analyses: One panelist remarked that “responsible companies” should be engaged in cost-benefit analyses,²⁵ while another panelist noted that some industry players are already performing risk-benefit analyses and discussions around methods to improve these type of analyses are ongoing.²⁶ We support further efforts in this area to determine which factors are needed to conduct effective analyses for improved data governance. Further, we suggest that use of cost-benefit analysis should not be limited to the private sector. For instance, The Internet Association respectfully encourages the Bureau of Consumer Protection’s Division of Privacy & Identity to Protection (DPIP) to work collaboratively with the Bureau of Economics and invite more economist to participate in them with respect to “big data.” This collaboration would offer further insight and a different perspective on how DPIP views the “big data” debate.
- Digital literacy programs: As noted in our pre-workshop comment submission, bottom-up solutions like education are critical to ensuring practical and effective privacy protections for users. Through educational initiatives, users can better understand how, when, and why their data is being used. Therefore, we urge the Commission to investigate effective methods for educating users in a “big data” context.
- Privacy research and development: In a recent multi-association letter, we respectfully encouraged the National Institute of Standards Technology “to focus its efforts on cataloguing, in a policy-neutral manner, how privacy engineers accomplish privacy-by-design or information management processes.”²⁷ The Internet Association continues to

²⁵ *Id.* at 238.

²⁶ *Id.* at 236.

²⁷ Letter from Application Developers Alliance, *et. al.*, to National Institute of Standards Technology (Oct. 10, 2014).



support the dedication of resources to supporting research and development that advance privacy-enhancing technologies such as de-identification.

We appreciate the opportunity to submit additional comments as the Commission continues to investigate “big data” and consumer protection issues, and look forward to further engagement in highlighting the myriad of benefits afforded by “big data” as this debate develops.

Respectfully submitted,

/s/ Michael Beckerman
Michael Beckerman
President & CEO
The Internet Association