**Opening Remarks FTC Chairwoman Edith Ramirez**
**Big Data: A Tool for Inclusion or Exclusion?**
**Washington, DC**
**September 15, 2014**

Welcome and thank you for joining us today. I also want to take this opportunity to

thank the FTC staff who organized today's program as well as our distinguished speakers for

sharing their expertise with us.

We are at a pivotal stage in the information age. Thanks to smartphones and smart

meters, wearable fitness devices, social media, connected cars, and retail loyalty cards, each of

us is generating data at an unprecedented rate. In fact, in 2013 it was reported that an astonishing

90 percent of the world's data was generated in the two preceding years.[1] Today, the output of

data is doubling every two years.[2] Advances in computational and statistical methods mean that

this mass of information can be examined to identify correlations, make predictions, draw

inferences, and glean new insights.

This is big data.

It has the capacity to save lives, improve education, enhance government services,

increase marketplace efficiency, and boost economic productivity. But the same analytic power

that makes it easier to predict the outbreak of a virus,[3] identify who is likely to suffer a heart

attack,[4] or improve the delivery of social services,[5] also has the capacity to reinforce

---

[1] *Big Data, for better or worse: 90% of world's data generated over last two years*, Science Daily (May 22, 2013), http://www.sciencedaily.com/releases/2013/05/130522085217.htm.

[2] Steve Lohr, *The Age of Big Data*, N.Y. TIMES (Feb. 11, 2012), *available at* http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?pagewanted=all.

[3] *How A Computer Algorithm Predicted West Africa's Ebola Outbreak Before It Was Announced*, Public Health Watch (Aug. 10, 2014), http://publichealthwatch.wordpress.com/2014/08/10/how-a-computer-algorithm-predicted-west-africas-ebola-outbreak-before-it-was-announced/.

[4] Aviva Rutkin, *Algorithm predicts heart attacks 4 hours before doctors, Big Data made simple (*Aug. 16, 2014), http://www.bigdata-madesimple.com/algorithm-predicts-heart-attacks-4-hours-before-doctors/.

disadvantages faced by low-income and underserved communities.  As businesses segment

consumers to determine what products are marketed to them, the prices they are charged, and the

level of customer service they receive, the worry is that existing disparities will be exacerbated.

Is this discrimination?  In one sense, yes.  By its nature, that is what big data does in the

commercial sphere – analyzes vast amounts of information to differentiate among us at lightning

speed through a complex and opaque process.  But is it unfair, biased, or even illegal

discrimination?  And if so, can steps be taken to level the playing field?

Those are the questions we will be exploring today.  Big data is at an early stage.  We

have the ability to shape its development and its outcomes.  If we are alert to the risks presented

by big data, we can take steps to guard against them.  We can help ensure that big data can be a

tool for economic inclusion, not exclusion.

That is the weighty subject before us today.  But before we begin the discussion, I would

like to address three questions.  First, how did we get here; second, what is our aim with today's

program; and, finally, where do we go from here?

## I.      How We Got Here

Let me start by tackling the first question – how did we get here – literally.  Whatever

mode of transportation you used to get to this workshop, there were apps or connected device

available to assist your commute.  Those of who came here using public transportation may have

availed yourselves of apps to tell you when the next bus or train would arrive.  If you came by

car, you may have benefitted from GPS technologies that gave you directions, sent you real-time

---

[5] Federico Guerrini, *From Homelessness to Human Trafficking, How A NYC Nonprofit is Using Big Data to Make a Difference*, FORBES (Apr. 4, 2014), *available at* http://www.forbes.com/sites/federicoguerrini/2014/04/24/from-syria-to-homelessness-and-human-trafficking-how-a-nyc-nonprofit-is-trying-to-make-a-difference-using-big-data/.

traffic alerts, or allowed you to summon a taxi or driver by tapping on a smartphone app. And for the virtuous among us who biked or walked here, you may have used a wearable device to track the distance traveled and calories burned. No matter your mode of transportation, once in the vicinity, an app or website may have helped you find a spot nearby to buy a cup of coffee before arriving at the workshop.

These very services and devices that helped many of us get here physically are also what brought us here figuratively. The popularity of smartphones and other mobile devices, the array of mobile apps we have at our fingertips, and the burgeoning Internet of Things more generally means that countless individuals actively and passively generate information in an extensive ecosystem throughout the day. The proliferation of connected devices, the plummeting cost of collecting, storing, and processing information, and the ability of data brokers and others to combine offline and online data mean that companies can accumulate virtually unlimited amounts of consumer information and store it indefinitely. Using predictive analytics, they can learn a surprising amount about each of us from this data.

While powerful algorithms can unlock value from immense datasets, their ability to draw correlations and make fine-grained distinctions also raises the prospect of differential treatment of low-income and underserved populations. This is a risk suggested by the Commission's recent report on the data broker industry.[6] The Commission's study of a cross-section of nine data brokers revealed that data brokers aggregate online and offline data from disparate sources to make inferences about consumers' ethnicity, income, religion, age, and health conditions,

---

[6] FED. TRADE COMM'N, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY (2014), *available at* http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

among other characteristics.  As the FTC and others have found, some brokers create segments

or clusters of consumers with high concentrations of minorities or low income individuals.[7]

There may be legitimate reasons why businesses would want to sort consumers in this fashion,

but the practice also raises the possibility that these segments will be used for what I have called

"discrimination by algorithm,"[8] and what others have called "digital redlining."[9]

We heard these concerns this past spring at the FTC's seminar on predictive scoring.

There are now products, beyond traditional credit scores, that purport to predict or "score"

everything from the chances that a transaction will result in fraud to the efficacy of sending

consumers catalogs and the best prices to offer consumers.  Some speakers lauded the benefits of

such predictions, emphasizing that they enable the personalization many consumers want and

help minimize the risk of fraud.

But other speakers worried that certain predictive scoring products could fall outside the

reach of the Fair Credit Reporting Act (FCRA) and the Equal Credit Opportunity Act (ECOA),

despite having an impact on consumers' access to credit, housing, employment, and insurance.

For example, if a company lowers my credit limit based on a score that reflects my credit history,

I would be entitled to certain protections under the FCRA.  If, however, the same company

---

[7] *See id.* at 20, 24, 47, 48, App. B-3; MAJORITY STAFF, OFFICE OF OVERSIGHT & INVESTIGATIONS, SENATE COMM. ON COMMERCE, SCIENCE, &TRANSP., A REVIEW OF THE DATA BROKER INDUSTRY:  COLLECTION AND SALE OF CONSUMER DATA FOR MARKETING PURPOSES 12, 24-26 (2013), *available at* http://www.commerce.senate.gov/public/?a=Files.Serve&File_id=bd5dad8b-a9e8-4fe9-a2a7-b17f4798ee5a; Charles Duhigg, *Bilking the Elderly with a Corporate Assist*, N.Y. TIMES (May 20, 2007), *available at* http://www.nytimes.com/2007/05/20/business/20tele.html?pagewanted=all.
[8] Edith Ramirez, *Protecting Consumer Privacy in a Big Data Age*, Address Before The Media Institute (May 8, 2014), *available at* http://www.ftc.gov/system/files/documents/public_statements/308421/140508mediainstitute.pdf.
[9] EXECUTIVE OFFICE OF THE PRESIDENT, BIG DATA:  SEIZING OPPORTUNITIES, PRESERVING VALUES 46, 53 (2014) [hereinafter White House Big Data Report], *available at* http://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.

lowers my credit limit based on the scores of a group in which I am a member, the application of the FCRA may be less clear. Will these scores be used in ways that influence the opportunities of low-income, minority, or other populations to get credit, jobs, housing, or insurance in ways that fall outside of the protections of the FCRA or ECOA? Could the use of geographic information, such as zip codes, for example, lead to Americans in low-income or rural neighborhoods being charged higher prices? And if so, is this a worrisome function of big data or just a continuation of age-old pricing practices and market forces?

These and other issues also figured prominently in the White House's wide-ranging report on big data, which squarely raised the concern that large-scale information analytics will be used for disparate or discriminatory outcomes for certain consumers, even absent discriminatory intent.[10]

## II.     Today's Program

It is questions and concerns raised by these prior initiatives that bring us to today's program, and to my second question – what is our goal with today's workshop? We will explore whether and how big data helps to include or exclude certain consumers from full opportunity in the marketplace. To help shed light on this issue, we have convened experts from industry, consumer and civil rights groups, academia, and government, representing a variety of perspectives. Our panelists and speakers will provide us a framework for our conversation today, assess current big data practices in the private sector, discuss possible developments on the horizon, present pertinent research, and offer potential ways to ensure that big data is a force for economic inclusion.

---

[10] *See* WHITE HOUSE BIG DATA REPORT at 7, 45-47, 51-53, 64.

It is my hope that our participants will discuss in depth the benefits and risks of big data to low-income and underserved populations.  On the benefits side, let me start the discussion with one example.  New York City is developing a tool that combines eviction data with emergency shelter admission information and other data to predict when individuals or families are on the brink of homelessness.  Using this information, the city is able to deploy social workers to help these families and prevent them from ending up on the street.  This is an example of positive government use, rather than a business use, but I hope our speakers will provide examples showing how companies can also use big data to benefit those in low-income or underserved groups.

As for real-world examples of possible risks, let me cite a study conducted by Latanya Sweeney, who is here from Harvard serving as the Commission's Chief Technologist.  Professor Sweeney found that web searches for distinctively black names were 25 percent more likely to produce an ad suggesting the person had an arrest record, regardless of whether that person had actually been arrested, than web searches for distinctively white names.[11]  This could have devastating consequences for job applicants and others, by creating the impression that the individual has been arrested.  While the research did not establish why the algorithm yielded these racially disparate results, it does provide a concrete example of how an algorithm may have adverse repercussions for a particular population.  I expect we will hear more illustrations today, including from Professor Sweeney who will be presenting results of a more recent study.

---

[11] Latanya Sweeney, *Discrimination in Online Ad Delivery* 56 COMMUNICATIONS OF THE ASS'N FOR COMPUTING MACHINERY 44, 44-54 (2013), *prepublication draft available at* http://dataprivacylab.org/projects/onlineads/1071-1.pdf.

**III.     Where Do We Go From Here?**

After we conclude our workshop, the question naturally arises, "where do we go from here?"  We may all have an array of apps to guide us home when we leave this afternoon, but there is no clear path for navigating the use of big data in a way that advances opportunities for all consumers while diminishing the risks of adverse differential impact on vulnerable populations.  We may not know what the best course should be, but I believe we should have at least three objectives going forward.

First, we should identify areas where big data practices might violate existing law.  Where they do, the FTC is committed to rigorous enforcement of the law, as demonstrated by cases such as our recent action against Instant Checkmate, a website that promoted some of its background checks as tools for screening tenants and employees.  The FTC alleged that Instant Checkmate did so without regard for the FCRA and obtained a $525,000 fine and a permanent injunction against the company.[12]  In addition to helping the FTC and others to enforce existing laws, today's program should also help identify any gaps in current law and ways to fill them.

Second, we need to build awareness of the potential for big data practices to have a detrimental impact on low-income and underserved populations.  I would like today's program to help foster a discussion about industry's ethical obligations as stewards of information detailing nearly every facet of consumers' lives.

---

[12] Complaint, United States v. Instant Checkmate, Inc., No. 3:14-cv-00675-H-JMA (S.D. Cal. Mar. 24, 2014); Stipulated Final Order, United States v. Instant Checkmate, Inc., No. 3:14-cv-00675-H-JMA (S.D. Cal. Apr. 1, 2014); *see also* Complaint, Equifax Info. Servs., LLC, No. C-4387 (F.T.C. Mar. 5, 2013), *available at* http://www.ftc.gov/sites/default/files/documents/cases/2013/03/130315equifaxcmpt.pdf; Complaint, United States v. Spokeo, Inc., No. CV12-05001 (C.D. Cal. June 7, 2012).

Third, and relatedly, we should encourage businesses to guard against bias or disparate impact on low-income and vulnerable populations when designing their analytics systems, algorithms, and predictive products. A good example is the Boston "Street Bump" app highlighted in the White House big data report.[13] Like any big city, Boston has its share of potholes and faces the ongoing challenge of staying on top of street repairs. To help address the issue, the city released a mobile app residents could use to identify potholes in need of repair. But the city also recognized that because lower-income individuals are less likely to carry smartphones, the data might skew road services to higher income neighborhoods. They addressed this problem by issuing the app to road inspectors who service all parts of the city equally and used the data gathered from the inspectors to supplement what they received from the public. This illustrates how consideration of risks before launching a product or service can help avoid them.

* * *

Big data can have big consequences. Those consequences can be either enormously beneficial to individuals and society or deeply detrimental. It will almost certainly be a mixture of the two. But it is the responsibility of the FTC and others to help ensure that we maximize the power of big data for its capacity for good while identifying and minimizing the risks it presents. As we navigate the transformative terrain of big data, it is vital that we work to ensure that technological innovation benefits all consumers, whatever their backgrounds. I look forward to hearing the thoughts and ideas of our panelists on how to do that. And I thank you all for your contributions to that endeavor.

---

[13] *See* WHITE HOUSE BIG DATA REPORT at 51-52.