Protecting Privacy in the Era of Big Data
Remarks of FTC Chairwoman Edith Ramirez
International Conference on Big Data from a Privacy Perspective
Hong Kong
June 10, 2015

I am delighted to be here in Hong Kong and want to thank Commissioner Allan Chiang for inviting me to participate in this conference to discuss the topic of big data and its impact on consumers and privacy, and to share what we are doing to address these issues at the U.S. Federal Trade Commission.

While big data itself is not new, we're in the midst of a new era – a big data revolution. And make no mistake, it's a game changer. Some argue it will transform "every sphere of life."[1] Without doubt, big data holds the promise of solutions to global problems – the potential to improve the quality of health care while cutting costs; enable forecasters to better predict the weather and spikes in energy consumption; and improve industrial efficiencies in order to deliver better and lower-cost products and services to consumers. It also has the potential to offer extraordinary, even life-altering, benefits for consumers that we can already see today – opening access to credit, increasing educational opportunities, and providing specialized healthcare.

But just as big data has the potential for big benefits, it also has the potential for big risks. As we are seeing in the United States, as companies develop new and innovative ways to "score" consumers, organizations can use these scores to deny consumers the ability to complete transactions, often without any explanation. Unscrupulous organizations can use big data to offer misleading offers or scams to the most vulnerable prospects, a trend that we at the U.S. Federal Trade Commission are unfortunately seeing today. Consumers themselves are wary. Just last

[1] The White House, *Big Data: Seizing Opportunities, Preserving Values* (May 2014), *available at* https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_5.1.14_final_print.pdf.

week, a study from the Annenberg School for Communication at the University of Pennsylvania concluded that many Americans do not think the trade-off of their data for personalized services, giveaways, or discounts is a fair deal.[2]

So how can we harness the transformative elements of the big data revolution while addressing the potential risks to consumer privacy? In my view, we have to address challenges head on through a multifaceted approach in order to reduce the potential risks.

What I would like to do this morning is to start by highlighting the life cycle of big data – how information is collected, compiled, analyzed, and used. I will then turn to the key privacy risks presented by big data, including a lack of transparency, loss of consumer control, heightened data security risks, and the potential for discrimination. Finally, I will offer proposed solutions for addressing these risks.

I. The Life Cycle of Big Data

In describing today's big data phenomenon, people often refer to the "three Vs" – volume, velocity, and variety.[3] The volume and variety of data both stem in part from the ubiquitous collection and compilation of smaller data – a tap of a smartphone, a website login, or a movement collected by a shopping center sensor, to cite just a few examples. The exponential increase in the volume of data of course also stems from the plummeting cost of data storage, which can extend the lifespan of data indefinitely. And the variety of today's big data reflects more than just the ever-increasing collection and storage of raw data. It is also the product of an

[2] *See* Joseph Turow, *et al*., *The Tradeoff Fallacy: How Marketers are Misrepresenting American Consumers and Opening Them Up to Exploitation* (June 2015) at 3-4.

[3] *See, e.g.*, Eric Savitz, *The New Factors of Production and the Rise of Data-Drive Applications*, FORBES (Oct. 31, 2011), *available at* http://www.forbes.com/sites/ciocentral/2011/10/31/the-new-factors-of-production-and-the-rise-of-data-driven-applications/.

unprecedented power to analyze data, in order to draw inferences about the world and make predictions about events to come.

As a first step to understanding the big data economy, we must understand the ever-growing sources of "little data" that are the building blocks of big data. Online tracking is one method of gathering little bits of data about who consumers are, what they do, and where they go. With every search, every click, every tweet, and every post, data is collected. Traditionally, companies have engaged in this type of tracking through cookies. And although browsers and self-regulatory organizations are providing consumers with tools to limit tracking through cookies, companies have been experimenting with other methods such as Flash cookies, history "sniffing," and device "fingerprinting" to track consumers across websites. On mobile devices, cross-app tracking is ubiquitous. Indeed, we have gone from cross-site tracking to cross-app tracking, and now, to cross-device tracking, where companies can track the same consumer across her desktop, laptop, tablet, and smartphone.

But data collection is not limited to computers and mobile devices. Companies also hope to follow consumers across the Internet of Things ("IoT"). Three and one-half billion sensors are already in the marketplace.[4] Today, connected devices are in our homes, our cars, and even on our bodies, in the form of connected smoke detectors and light bulbs, connected cars, and wearable computers, among others. This is only the beginning. Some experts estimate there will be 25 billion connected devices by year end and 50 billion by 2020.[5] Others view those estimates as conservative, predicting the number of sensors will increase to trillions within the

[4] *See* Stanford University, *TSensors Summit™ for Trillion Sensor Roadmap* (Oct. 23-25, 2013), *available at* http://tsensorssummit.org/Resources/Why%20TSensors%20Roadmap.pdf.

[5] Dave Evans, Cisco Internet Business Solutions Group, *The Internet of Things: How the Next Evolution of the Internet is Changing Everything* (2011) at 3, *available at*

next decade.[6] All of these connected devices mean much more data. Globally, mobile data traffic will exceed fifteen exabytes – about 15 quintillion bytes – each month.[7] By comparison, according to one estimate, a transcript of every word ever spoken would consume about five exabytes.[8]

And finally, though many big data discussions focus on these high-tech methods of data collection, we should not forget good old-fashioned brick-and-mortar data collection. Through loyalty programs, warranty cards, surveys, sweepstakes entries, and even credit card purchases, we leave yet another trail of "little data" bread crumbs.

After collection, the next step in the life cycle of big data is compilation and consolidation. Entities that compile data may include online ad networks, social media companies, and large banks or retailers. But perhaps the largest category of entities that compile and consolidate data are data brokers, which combine "little data" from disparate sources to build profiles about individual consumers. Last year, the FTC released a report studying how data brokers acquire and store billions of data elements on nearly every U.S. consumer.[9] In our report, we also discussed the complexity of the data broker industry, in which multiple layers of data brokers provide data to each other, making it virtually impossible for consumers to determine where their data originated.

---

http://www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf. These estimates include all types of connected devices, not just those aimed at the consumer market.

[6] *Id*.

[7] Cisco, *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018* (2014) at 3, *available at* http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf.

[8] BBC News, *Does Online Video Threaten the Net?* (Apr. 29, 2008), *available at* http://news.bbc.co.uk/2/hi/technology/7370956.stm.

[9] FTC Report, *Data Brokers: A Call for Transparency and Accountability* (May 2014), *available at* https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf (hereinafter "Data Broker report").

The third step is analytics – the use of data science to make predictions about people, including drawing potentially sensitive inferences. For example, in our data broker report, we described how data brokers use information they obtain to put consumers into categories. Some of these seem relatively benign, such as "Dog Owner," "Winter Activity Enthusiast," or "Mail Order Responder."[10] Other categories involve more sensitive inferences. This is true of categories that primarily focus on consumers' ethnicity and income levels, such as "Urban Scramble" and "Mobile Mixers," both of which include a high concentration of Latinos and African Americans with low incomes.[11] Other potentially sensitive categories include combined inferences about a consumer's age and economic status, such as the "Thrifty Elders" category or a category called "Rural Everlasting," which includes single men and women over the age of 66 with "low educational attainment and low net worth."[12] Yet other categories highlight potentially sensitive health-related topics or conditions, such as "Expectant Parent," "Diabetes Interest," and "Cholesterol Focus."[13]

The final step in the life cycle of big data is use. While data brokers compile and analyze big data, they ultimately sell their products to marketers, retailers, banks, governments, and educational institutions. And how is the data used? Take the IoT. Connected cars may direct emergency responders to an accident, but will the data transmitted be shared with your insurer who may raise your rate or cancel your policy? Your smart TV may track your favorite shows, but will your TV-viewing habits be shared with prospective employers or schools?

When it comes to the use of big data, we also must consider the potential for illegal or discriminatory uses. For example, last fall the FTC held a workshop entitled *Big Data: A Tool*

[10] *Id.* at 47.
[11] *Id.*
[12] *Id.*

for Inclusion or Exclusion?[14] In this workshop, we specifically examined the impact on low-income and underserved consumers when big data is used to make inferences and predictions about them. We learned that, among the many benefits to underserved consumers, big data can increase their access to credit, by enabling alternative credit scores for consumers who lack traditional credit histories and previously were considered ineligible for credit. Big data can also increase access to education, for example, by identifying students for advanced classes who would have been excluded by the usual selection criteria or students who are at risk of dropping out and need additional help. And it can improve access to health care, by identifying individuals most at risk of illness or hospitalization.

But workshop participants and commenters also emphasized the potentially harmful effects when big data is used to draw inferences about consumers. Just as new scoring models can be used to extend credit, education, and health resources to underserved consumers, these models pose the risk of what others and I have called "discrimination by algorithm." This occurs when facially neutral algorithms discriminate against or adversely impact low-income and economically vulnerable consumers. For instance, if online companies charge consumers in different neighborhoods different prices, one result could be that consumers in poorer areas pay more for online products than those in affluent communities. If educational institutions use analytics to identify elementary school students who are not likely to go to college, these students may never have access to information about college-prep courses or financial aid packages.

---

[13] *Id.*

[14] *See* FTC Workshop, *Big Data: A Tool for Inclusion or Exclusion?* (Sept. 15, 2014), *available at* https://www.ftc.gov/news-events/events-calendar/2014/09/big-data-tool-inclusion-or-exclusion.

II. Key Risks

As you can see, every step in the life cycle of big data raises the potential for significant privacy and other risks. First, despite the potential for big data's positive impact on the lives of consumers, there is a real risk of lack of transparency and loss of consumer control. Consumers would likely be surprised to know the myriad of ways in which data is collected about them. And virtually all of the data broker activities described in our report, including the placement of consumers into potentially sensitive categories, take place without consumers' knowledge or control.[15] Furthermore, consumers' ability to access and correct the information that data brokers hold about them is limited, where it exists at all.[16] As a result, there is a risk that if a consumer is denied the ability to complete a transaction based on inaccuracies in a data broker's profile, the consumer will be harmed without knowing why and without being able to address the problem.

Second, there is a risk of unexpected and unwelcome use of data. For example, data generated by connected medical devices could be used to make credit, insurance, and employment decisions without consumers' knowledge or consent, and without ensuring the accuracy of the data. Big data could also be used in ways that could exacerbate existing socio-economic disparities, by segmenting consumers with regard to the customer services they receive, the prices they are charged, and the types of products that are marketed to them.[17]

Third, big data raises concerns about data security. For instance, by compromising IoT devices, hackers could gain access to the same types of sensitive financial account information,

[15] *See Data Broker report, supra* n.11, at 48.
[16] *Id.*
[17] *Id.*

passwords, and other information used to commit identity theft or fraud.[18] Hackers might also exploit security vulnerabilities in devices such as smart cars or connected medical devices to create risks to physical safety in some cases. These potential risks are exacerbated by the fact that some companies entering the IoT market may not be as focused on security issues as those who have been manufacturing computer hardware and software for decades. And some companies may not even offer patching or security updates since some IoT devices may be inexpensive and essentially disposable.[19]

III. Solutions

As the primary U.S. agency charged with protecting consumer privacy in the commercial sphere, the FTC seeks to address many of these risks through law enforcement. The main statute we enforce is Section 5 of the FTC Act, which prohibits "unfair or deceptive acts or practices."[20] Under Section 5, we have brought actions against companies that have violated promises to refrain from sharing data with third parties[21] or to provide consumers choices about sharing.[22] We have settled 53 actions against companies that failed to maintain reasonable security of consumer data.[23]

---

[18] *See* FTC Staff Report, *Internet of Things: Privacy and Security in a Connected World* (Jan. 2015) at 10-12, *available at* https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf.

[19] *Id.* at 13-14.

[20] 15 U.S.C. § 45(a).

[21] *See, e.g.*, *FTC v. Goldenshores Technologies, LLC*, No. C-4446 (F.T.C. Dec. 5, 2013) (consent order); *FTC v. Myspace LLC*, No. C-4369 (F.T.C. May 8, 2012) (consent order).

[22] *See, e.g., United States v. Path, Inc.*, No. C-13-0448 (N.D. Cal. Feb. 1, 2013); *FTC v. Compete, Inc.*, No. C-4384 (F.T.C. Oct. 22, 2012) (consent order); *FTC v. Facebook, Inc.*, No. C-4365 (F.T.C. Nov. 29, 2011) (consent order); *FTC v. Google Inc.*, No. C-4336 (F.T.C. Mar. 30, 2011) (consent order); *FTC v. Chitika, Inc.*, No. C-4324 (F.T.C. Mar. 14, 2011) (consent order).

[23] *See, e.g.*, *FTC v. Snapchat, Inc.*, No. C-4501 (F.T.C. May 14, 2014) (consent order); *FTC v. Fandango, LLC*, No. C-4481 (F.T.C. Mar. 28, 2014) (consent order); *FTC v. Credit Karma, Inc.*, No. C-4480 (F.T.C. Mar. 28, 2014) (consent order); *FTC v. Twitter, Inc.*, No. C-4316 (F.T.C. June 25, 2010) (consent order); *FTC v. Reed Elsevier Inc.*, No. C-4226 (F.T.C. Mar. 27, 2008) (consent order).

We have also addressed issues involving the IoT and data brokers. For example, our TRENDnet settlement – the FTC's first Internet of Things case – involved a video camera designed to allow consumers to monitor their homes remotely for purposes ranging from home security to baby monitoring.[24] We alleged that although TRENDnet claimed that its cameras were "secure," it had faulty software that left the cameras vulnerable to online viewing, and in some instances listening, by anyone with the cameras' Internet address. This resulted in hackers posting 700 consumers' live feeds on the Internet.

And in our case against data broker LeapLab, we alleged that the company bought the payday loan applications of financially strapped consumers and then sold this sensitive information to marketers whom it knew had no legitimate need for it.[25] This included phony internet merchants that used the information to withdraw millions of dollars from consumers' accounts without their authorization. We charged that LeapLab's sale of this data to scam artists and others with no legitimate need for it is an unfair practice under the FTC Act.

We have also been a primary enforcer of one of the first U.S. laws to specifically address big data practices: the Fair Credit Reporting Act.[26] Where a data broker sells consumers' information to entities making decisions about credit, employment, housing, or other benefits, the FCRA may apply. If a company buys information about a consumer from a third party, and uses that information to deny a consumer credit, employment, housing, or other benefits, the company must notify consumers and give them an opportunity to correct inaccurate information. We have brought over 100 cases alleging violations of the FCRA. The message from our FCRA cases is

[24] *TRENDnet, Inc.*, No. C-4426 (Jan. 16, 2014) (consent), *available at* http://www.ftc.gov/enforcement/cases-proceedings/122-3090/trendnet-inc-matter.

[25] *FTC v. Sitesearch Corp. d/b/a LeapLab, LLC et al.*, FTC File No. 142-3192 (D. Az. filed Dec. 22, 2014), *available at* https://www.ftc.gov/enforcement/cases-proceedings/142-3192/sitesearch-corporation-doing-business-leaplab.

that if companies buy data about consumers from analytics companies and data brokers, and use that data to make eligibility determinations about consumers, the FCRA may apply.

Finally, we enforce the Equal Credit Opportunity Act, which prohibits discrimination in lending based on protected characteristics such as race, gender, and age.[27] If a company uses big data in a way that denies protected classes access to credit, ECOA may apply.

Despite our law enforcement efforts, we recognize that the laws we enforce are far from a perfect fit for today's marketplace and may have significant gaps. For example, the FCRA does not apply when businesses use their own in-house data analytics to make decisions about their customers or employees. And although ECOA would prohibit racial distinctions in terms of access to credit, it may not prohibit those distinctions in the types of advertisements served. Thus, a minority consumer may only see ads for subprime products and may never know about the availability of better credit offers. Finally, Section 5 does not require notices or choices about big data practices. This is why our efforts at the FTC must go beyond enforcement of existing laws. I will continue to urge our Congress to enact comprehensive privacy and data security legislation. I also believe businesses should take at least the following three steps.

First, all entities that have a role in the big data life cycle must step up their efforts to provide consumers with transparency and choice. This includes the entities that collect, compile, consolidate, and analyze data, and the entities that use big data. We have brought several cases that stand for this proposition. For example, we recently took action against an analytics company that tracked consumers' mobile devices in retail stores.[28] We alleged that the company, Nomi Technologies, offered consumers two ways to exercise choices not to be tracked, but failed

[26] 15 U.S.C. § 1681e.
[27] 15 U.S.C. § 1691(a).

to provide one of those choices. In another context, we took action against a popular flashlight app for failing to disclose to consumers that it was providing geolocation information to ad networks.[29] The message from these cases is that, regardless of where you are in the data chain, you have responsibilities with respect to transparency and choice.

Second, companies need to take reasonable steps to safeguard consumers' personal information. The importance of implementing reasonable data security cannot be overstated. In addition to the 53 settlements I mentioned, we also aim to educate businesses on how to implement reasonable security. We have a host of materials on our website, both general tips for businesses as well as tips in specific areas such as mobile and the IoT. And we recently announced an initiative called Start with Security, which will include new educational materials, webinars, and workshops aimed at giving businesses practical advice on security. Our first workshop will take place in San Francisco on September 9.

Finally, businesses must use big data responsibly, in ways that do not discriminate against or adversely affect vulnerable populations. At our big data workshop, for instance, panelists discussed the fact that big data analytics may show that people with longer commuting distances are more likely to leave a job. If organizations screen job applicants based on commuting distance, this could have a disparate impact on minority communities that often have to travel long distances to commercial hubs where most jobs are located. Responsible companies have chosen not to include this type of analytics in their hiring decisions.

Companies should think carefully about how the data sets they use have been crafted. If they identify potential biases in the creation of these data sets, companies can and should develop

---

[28] *See Nomi Technologies, Inc.*, FTC File No. 132-3251 (F.T.C. Apr. 23, 2015) (proposed consent order), *available at* https://www.ftc.gov/enforcement/cases-proceedings/132-3251/nomi-technologies-inc-matter.

strategies to overcome them.

IV. Conclusion

At the FTC, we will continue to shine a light on these issues. I believe that all of us must continue to learn and raise public awareness about the ways that big data is collected and used. And we must find ways to make big data decision-making about consumers more transparent. It is only through such knowledge that we can begin to address the potential privacy and other risks while continuing to reap the benefits of big data.

Thank you.

[29] *See FTC v. Goldenshores Technologies, LLC*, No. C-4446 (F.T.C. Dec. 5, 2013) (consent order), *available at* https://www.ftc.gov/enforcement/cases-proceedings/132-3087/goldenshores-technologies-llc-erik-m-geidl-matter.