THE PRIVACY CHALLENGES OF BIG DATA: A VIEW FROM THE LIFEGUARD'S CHAIR

Keynote Address by FTC Chairwoman Edith Ramirez
(As Prepared for Delivery)
Technology Policy Institute Aspen Forum
Aspen, Colorado
August 19, 2013

I want to thank all of you for coming this afternoon. The temptation to skip this talk and hike up to the Maroon Bells must have been nearly overwhelming. I am glad that you did not succumb, though I would have understood. As Oscar Wilde once quipped, "I can resist anything except temptation."

My topic today is "big data" and the privacy challenges it may pose to consumers. I want to explore how we can reap the benefits of big data without falling prey to possible pitfalls.

There is little doubt that the skyrocketing ability of business to store and analyze vast quantities of data will bring about seismic changes in ways that may be unimaginable today. Unlocking the potential of big data can, for instance, improve the quality of health care while cutting costs. It can enable forecasters to make increasingly precise predictions about weather, crop yields, and spikes in the consumption of electricity. And big data can improve industrial efficiency, helping to deliver better products and services to consumers at lower costs.

But one might ask: Can the Federal Trade Commission or any governmental entity safeguard consumer privacy given the breakneck pace of technological innovation? In my view, the answer is "yes." The fact that "big data" may be transformative does not mean that the challenges it poses are, as some claim, novel or beyond the ability of our legal institutions to respond. Take the beautiful leaves of the Aspen trees. In late summer, they turn from green to bright yellow and the forest takes on a golden glow. That transformation is a breathtaking one; it is one of the wonders of the West. But not everything breathtaking is new. The transformation of the Aspen trees happens every year.

The emergence of big data is similarly breathtaking and potentially game changing. But the challenges it poses to consumer privacy are familiar, even though they may be of a magnitude we have yet to see. The solutions are also familiar. And, with the advent of big data, they are now more important than ever. Addressing the privacy challenges of big data is first and foremost the responsibility of those collecting and using consumer information. The time has come for businesses to move their data collection and use practices out of the shadows and into the sunlight. But the FTC has a critical role to play as well. This afternoon, I will address both how businesses should approach big data to protect consumer privacy and how the FTC will work to ensure companies live up to that obligation.

I. THE FTC'S ROLE IN BIG DATA

Let me begin with my vision of the FTC and its role in light of the emergence of big data. I grew up in a beach town in Southern California. To me, the FTC is like the lifeguard on a beach. Like a vigilant lifeguard, the FTC's job is not to spoil anyone's fun but to make sure that no one gets hurt. With big data, the FTC's job is to get out of the way of innovation while making sure that consumer privacy is respected.

Congress gave us several tools to do just that. Under the FTC Act, it tasked the Commission with preventing unfair or deceptive acts or practices that may affect interstate commerce. This mandate gives the FTC authority over deceptive claims about matters that are important to consumers, including privacy and data security. For instance, in the FTC's actions against Google, Facebook, Myspace and others, we alleged that each of these companies deceived consumers by breaching commitments to keep their data confidential.[1] That isn't OK, and it is the FTC's responsibility to make sure that companies live up to their commitments.

Congress also assigned the FTC the responsibility to prevent "unfair" commercial practices — that is, conduct that substantially harms consumers, or threatens to substantially harm consumers, which consumers cannot reasonably avoid, and where the harm outweighs the benefits. The FTC has used its unfairness authority against companies that fail to provide reasonable data security. Take one example: Last year, we sued the Wyndham hotel chain for poor data security practices that led to three data breaches in an 18-month period.[2] Over a half-million credit card files ended up in the hands of an identity-theft ring operating through domains registered in Russia. All told, the FTC has brought over 40 data security cases under our unfairness and deception authority, many against very large data companies, including LexisNexis,[3] ChoicePoint,[4] and Twitter,[5] for failing to provide reasonable security safeguards.

---

[1] *United States v. Google, Inc.*, No. 5:12-cv-04177 (N.D. Cal. filed Aug. 8, 2012), *available at* http://www.ftc.gov/os/caselist/c4336/120809googlecmptexhibits.pdf; *Google, Inc.*, No. C-4336 (F.T.C. Oct. 24 2011), *available at* http://www.ftc.gov/os/caselist/1023136/111024googlebuzzcmpt.pdf; *Facebook, Inc.*, No. C-4365 (F.T.C. Aug. 10, 2012), *available at* http://www.ftc.gov/os/caselist/0923184/120810facebookcmpt.pdf; *Myspace, Inc.*, No. C-4369 (F.T.C. Sept. 11, 2012), *available at* http://ftc.gov/os/caselist/1023058/120911myspacecmpt.pdf.

[2] *FTC v. Wyndham Worldwide Corp.*, No. CV 12-1365 (D. Ariz. filed Jun 26, 2012), *available at* http://www.ftc.gov/os/caselist/1023142/120809wyndhamcmpt.pdf.

[3] *Reed Elseveir, Inc.*, No. C-4226 (F.T.C. Aug. 1, 2008), *available at* http://www.ftc.gov/os/caselist/0523094/080801reedcomplaint.pdf.

[4] *United States v. ChoicePoint, Inc.*, No. 1:06-cv-0198 (N.D. Ga. Feb. 15, 2006), *available at* http://www.ftc.gov/os/caselist/choicepoint/0523069complaint.pdf..

[5] *Twitter, Inc.*, No. C-4316 (F.T.C. Mar. 11, 2011), *available at* http://www.ftc.gov/os/caselist/0923093/110311twitterdo.pdf.

Congress has also charged the FTC with enforcing a number of sector-specific privacy laws, including the Fair Credit Reporting Act or "FCRA," which can be seen as the first "big data" privacy law. Congress enacted the FCRA because it was worried that growing databases — the "big data" of the 1970s — could be used in ways that were invisible and harmful to consumers.

The FCRA sets out rules for companies that use data to determine creditworthiness, insurance eligibility, suitability for employment, and to screen tenants. When companies obtain this data from a credit bureau or other consumer reporting agency, they must notify consumers when the information contributed to decisions that adversely affect them — for example, denying them access to credit or giving them less-than-favorable credit terms, or turning them down for a job or insurance.

Another statute the FTC enforces is the Children's Online Privacy Protection Act, or "COPPA," which requires companies to get a parent's consent before collecting personal information from kids under 13. We recently updated our rule implementing COPPA to respond to collection practices made possible by new technology, namely, data-gathering tools like social media and mobile applications.

So let's turn to the key big data questions. The term "big data" refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze.[6] And it's no surprise that datasets are expanding exponentially. Under "Moore's Law," the ability to store, aggregate and combine data and conduct deep analyses has skyrocketed.[7] McKinsey reports that in 2011, a consumer could purchase a disk to store all of the world's music for under $600.[8] If Moore is right, the cost today would be less than $300.

This phenomenal growth in storage and analytic power means that big data is no longer the province of a few giant companies, like large data brokers, banks, insurers, and health care providers. Big data is now, or soon will become, a tool available to all sectors of the economy.

Of course, many uses of big data bring tangible benefits to consumers and businesses alike. And many uses of big data raise no threats to consumer privacy. For example, many firms use big data analytics for purposes that have nothing to do with individuals — forecasting weather and stock and commodity prices; upgrading network security systems; and improving manufacturing supply chains.

---

[6] *See* MCKINSEY & CO., BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION AND PRODUCTIVITY 1 (June 2011), *available at* http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

[7] *See id.* at 2 & n.3.

[8] *Id.* at 2.

On the other hand, many firms use big data in ways that implicate individual privacy. That data may reflect an individual's health concerns, browsing history, purchasing habits, social, religious and political preferences, financial data and more. They may do so in the service of innovation and efficiencies that confer substantial benefits on consumers. As I have said, the FTC's role isn't to stand in the way of innovation; it is to ensure that these advances are accompanied by sufficiently rigorous privacy safeguards.

II. ADDRESSING THE RISKS OF BIG DATA

Earlier I mentioned the possible pitfalls associated with big data. Let me address the hazards we must avoid:

A. Indiscriminate collection of data

One risk is that the lure of "big data" leads to the indiscriminate collection of personal information. Some big data proponents argue that data is now the raw material of innovation, and therefore more data is always better. We are told that during the Industrial Revolution, there was no such thing as too much coal and iron ore. The resulting steel sparked the innovation that transformed the world — skyscrapers, high speed trains, and so on. Today's raw material, the argument goes, is data, and we need as much of it as we can collect.

That's a bridge — maybe even a steel bridge — I wouldn't buy. The indiscriminate collection of data violates the First Commandment of data hygiene: Thou shall not collect and hold onto personal information unnecessary to an identified purpose. Keeping data on the off-chance that it might prove useful is not consistent with privacy best practices. And remember, not all data is created equally. Just as there is low quality iron ore and coal, there is low quality, unreliable data. And old data is of little value. Is there really any worth to my law school search history when I was struggling to understand the rule against perpetuities? Should that data be held in perpetuity?

B. The Need to Ensure Meaningful Consumer Choice

A related concern is that some big data advocates insist that, because more data is always better, and because providing consumer choice may be especially challenging when it comes to big data, the time has come to reconsider limits on data collection. They contend that, to the extent that privacy protection is needed, the focus should be on after-the-fact *use* restrictions, not on limiting *collection*.

That argument stands privacy protection on its head. Let's go back to basics. Big data doesn't start as big data. Rather it is assembled, bit-by-bit, from little data and becomes "big" only when compiled into enormous databases. The little data often reflects deeply personal information about individuals: the medical treatment they receive; the products and services they buy; their physical location; the websites they surf; their intimate communications with family

and friends; and the list goes on.

Personal information should not be collected against consumers' wishes.[9] Business leaders understand this. They know that the quickest way to squander consumer trust is to go behind consumers' backs when collecting and using personal data.

Adding to the concerns is the reality that some collection of personal data is not, in fact, authorized by consumers. Consumers do, of course, often decide to share personal data. They post personal information on social networks. And in many instances they consent to the collection of personal data by businesses and service providers. But that consent is generally limited to the transaction at hand — for example, to enable lenders to evaluate mortgage applications or companies to ship items purchased on the Internet. Rarely, if ever, are consumers given a say about the aggregation of their personal data or secondary uses that are not even contemplated when their data is first collected.

The already intricate data-collection ecosystem is becoming even more complex. For example, the Internet of Things means more parts of our daily lives will generate data. Households with "smart" appliances such as refrigerators, televisions, thermostats — all saving energy and providing services well beyond the capabilities of "dumb" appliances — will soon be widespread. These devices will be connected to the Internet, collecting information that will end up in the hands of manufacturers, service providers and others. What are the privacy and security implications? These are questions we are thinking about at the FTC, and we will hold a workshop in November to explore the unique issues associated with smart technology.

Let me be clear, though. Focusing on consumer choice at the time of collection is critical, but use restrictions have their place too. No system of consumer choice can be perfect and some uses of data are clearly out-of-bounds. The Fair Credit Reporting Act, which I mentioned earlier, is a classic use restriction. Other statutes also impose restrictions on the use of personal data — from genetic information to details about campaign donations.[10]

---

[9] As the Commission explained last year in its Privacy Report, businesses should provide simplified choice before collecting consumer data for practices that are inconsistent with the context of the transaction or the company's relationship with the consumer, unless specifically authorized or required by law. *See* FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS 27 (Mar. 2012) [hereinafter FTC PRIVACY REPORT], *available at* http://www.ftc.gov/os/2012/03/120326privacyreport.pdf.

[10] *See, e.g.*, Genetic Information Nondiscrimination Act, 42 U.S.C. § 2000ff *et seq*. (limiting the disclosure and use of genetic information in health insurance and employment); Federal Election Campaign Act, 2 U.S.C. § 438(a)(4) (forbidding commercial use of campaign donation information filed with the Federal Election Commission made available for public inspection); Driver's Privacy Protection Act, 18 U.S.C. § 2721 *et seq*. (placing limits on the commercial use of drivers' license information obtained from state motor vehicle agencies).

As important as they are, use restrictions have serious limitations and cannot, by themselves, provide effective privacy protection. Information that is not collected in the first place can't be misused. And enforcement of use restrictions provides little solace to consumers whose personal information has been improperly revealed. There's no putting the genie back in the bottle.

Use restrictions are also hard to enforce. It is often difficult to identify the culprit when data is misused. Look at your smart phone: It contains a wealth of personal information, such as contacts, photographs, geo-location data, emails and incoming and outgoing phone numbers.

Multiple players in the mobile ecosystem have access to that data — not just the carriers that provide service, but the mobile applications you use, the advertising networks that service those applications, the analytics companies that receive the information from the networks, and many others.

If data is stripped from your smartphone and misused, it may be impossible to determine which one of many entities is responsible — an endeavor made more difficult by the archaic division of responsibility between the Federal Communications Commission, which has jurisdiction over carriers, and the FTC, which, in many instances, has jurisdiction over all of the players *except* the carriers. The FTC has long urged Congress to rectify this ancient allocation by repealing the "common carrier" exception to the FTC's jurisdiction,[11] and the rise of mobile technology underscores the need for Congressional action.

C. Data Breach

The risk of a data breach is also not trivial. After all, the larger the concentration of sensitive personal data, the more attractive a database is to criminals, both inside and outside a firm. And the risk of consumer injury increases as the volume and sensitivity of the data grows.

An analogy makes my point: If water in a glass spills, the consequences generally are manageable. One wipes up the water and moves on. But if one builds a dam to store tremendous volumes of water and the dam breaks, the consequences can be quite serious.

In other words, with big data comes big responsibility. Firms that acquire and maintain large sets of consumer data must be responsible stewards of that information. The FTC can already bring actions under Section 5 of the FTC Act, and we will continue to be active in data security under my watch. But stronger incentives to push firms to safeguard big data must be in place. The FTC has urged Congress to give the agency civil penalty authority against companies that fail to maintain reasonable security.[12] The advent of big data only bolsters the need for this

[11] 15 U.S.C. § 45(a)(2).

[12] *See, e.g.*, FTC PRIVACY REPORT, *supra* note 9, at 12 & n.65.

legislation.

D. Behind-the-Scenes Profiling

Firms of all sorts are using consumer data in ways that may not just be contrary to consumers' expectation, but could also be harmful to their interests. This problem is perhaps seen most acutely with data brokers — companies that collect and aggregate consumer information from a wide array of sources to create detailed profiles of individuals. Their success depends on having more and better data than their rivals. The concern is that their mega-databases may contain highly sensitive information. The risk of improper disclosure of sensitive information is heightened because consumers know nothing about these companies and their practices are invisible to consumers.

Last year, in our Privacy Report, the FTC called on data brokers to give consumers access to their information through an easy-to-find, easy-to-use common portal. The agency also supported legislation to give consumers access to, and a right to dispute or suppress, data held by brokers.[13]

To drill down even deeper into the workings of the industry, the Commission has also issued subpoenas — what the FTC calls 6(b) orders — to nine data brokers. The orders seek information about the nature and sources of the consumer information the data brokers collect; how they use, maintain, and disseminate the information; and the extent to which they allow consumers to access and correct their information or opt out of having their personal information sold.[14] We expect to issue a report later this year with our findings.

E. Data Determinism

The involuntary revelation of sensitive personal information is an important concern but it is a risk that predates big data and is inherent in the collection and use of personal information. There is another risk that *is* a by-product of big data analytics, namely, that big data will be used to make determinations about individuals, not based on concrete facts, but on inferences or correlations that may be unwarranted.

Let's call this possibility "data determinism." Individuals may be judged not because of what they've done, or what they will do in the future, but because inferences or correlations drawn by algorithms suggest they may behave in ways that make them poor credit or insurance risks, unsuitable candidates for employment or admission to schools or other institutions, or unlikely to carry out certain functions.

---

[13] *Id.* at 14, 68-70.

[14] *See* Fed. Trade Comm'n, Press Release, *FTC to Study Data Broker Industry's Collection and Use of Consumer Data* (Dec. 18, 2012), http://www.ftc.gov/opa/2012/12/databrokers.shtm.

Using correlations in this reductive way may make sense to the companies that rely on them. The law of big numbers means that, if the algorithm is correct, in general the companies will be making the right call. An error rate of one-in-ten, or one-in-a-hundred, may be tolerable to the company. To the consumer who has been mis-categorized, however, that categorization may feel like arbitrariness-by-algorithm.

The fact that decision-by-algorithm may be less than perfect is not to condemn the enterprise. Far from it. Using data-driven analytics to improve decision-making may be an important step forward. After all, human decision-making is not error-free. People often make imperfect decisions for a variety of reasons, including incomplete information, poor decisional tools, or irrational bias. But the built-in imperfections in the decision-by-algorithm process demand transparency, meaningful oversight and procedures to remediate decisions that adversely affect individuals who have been wrongly categorized by correlation. At the very least, companies must ensure that by using big data algorithms they are not accidently classifying people based on categories that society has decided — by law or ethics — not to use, such as race, ethnic background, gender, and sexual orientation.

Of course, where the Fair Credit Reporting Act applies, consumers already have access and correction rights. The FTC has been stepping up its FCRA enforcement efforts,[15] and that trend will continue.

III. STEPS BUSINESSES CAN TAKE TO HARNESS THE POWER OF BIG DATA WHILE SAFEGUARDING CONSUMERS

So far, I have talked about the possible hazards of big data, and that's in keeping with the FTC's role here: to worry about the potential adverse impacts of the use of big data and to figure out how to mitigate those risks. The FTC's privacy agenda aims to persuade companies to minimize risks in ways that encourage, not undercut, their ability to reap the rewards of a data-driven economy. The FTC urges companies to follow the three core principles laid out in the FTC's 2012 Privacy Report: privacy-by-design, simplified choice, and greater transparency. We

[15] *See, e.g., United States v. Certegy Check Servs., Inc.*, No. 1:13-cv-01247 (D.D.C. filed Aug. 15, 2013) (consent decree imposing $3.5 million fine for alleged FCRA violations by firm providing check authorization services); *Filiquarian Publ'g, LLC*, No. C-4401 (F.T.C. May 1, 2013) (alleging FCRA violations by provider of criminal background check service and approving consent order imposing injunction); *Equifax Info. Servs. LLC*, No C-4387 (F.T.C. Mar. 15, 2013) (alleging violations of the FCRA and FTC Act for selling lists of consumers who were late on mortgage payments and approving consent order requiring disgorgement of gross revenues from such sale); *United States v. Direct Lending Source, Inc.*, No. 12-CV-2441 (S.D. Cal. filed Oct. 12, 2012) (consent order imposing $1.2 million fine for alleged violations of FCRA in purchasing lists of consumers who were late on mortgage payments from Equifax); *United States v. Spokeo, Inc.*, No. CV12-05001 (C.D. Cal. filed June 7, 2012) (consent decree imposing $800,000 fine for alleged FCRA violations by data broker that compiled and sold personal information for screening job applicants).

have also urged Congress to enact baseline privacy legislation that is informed by these principles,[16] and I renew that call today.

A. Privacy by Design

Privacy by design means building privacy in as products and services are being developed. To do that, companies need to perform risk assessments to lay bare vulnerabilities by asking tough questions: Are security measures appropriate given the volume and sensitivity of the data? Is the analytic inquiry sufficiently fine-tuned to draw correlations with reasonable confidence? Could the contemplated uses cause harm to individuals, such as financial loss, injury to reputation, or unlawful discrimination? And if so, can these risks be avoided or mitigated? Once risks have been assessed, privacy by design enables prudent engineers to design systems that better safeguard privacy.

B. Simplified Choice

We also need to go back to first principles and take consumer choice seriously. Too often the "notice" part of this process is overlooked, even though it is a prerequisite to meaningful choice. Consumers must be told who is collecting their data and what the data will be used for. And choice mechanisms must be simple and easy-to-use. We need to design data acquisition methodologies that provide transparency from the very first interaction, and reaffirm transparency throughout the lifecycle of personal data. To that end, in the context of the collection of data that occurs as consumers surf the web, we need a workable Do Not Track option that will put consumers back in control over the collection of their information.

C. Transparency

But the need for greater transparency is not limited to collection of data about consumers' online behavior. A recurring theme I have emphasized — and one that runs through the agency's privacy work — is the need to move commercial data practices into the sunlight. For too long, the way personal information is collected and used has been at best an enigma "enshrouded in considerable smog."[17] We need to clear the air.

---

[16] FTC PRIVACY REPORT, *supra* note 9, at 13; Prepared Statement of the Federal Trade Commission Before the U.S. Senate Committee on Commerce, Science and Transportation, *The Need For Privacy Protections: Perspectives From the Administration and the Federal Trade Commission* 4-5 (May 9, 2012), *available at* http://ftc.gov/os/testimony/120509privacyprotections.pdf.

[17] This phrase is borrowed from judicial decisions describing imponderable questions. *See, e.g.*, *General Motors Corp. v. Ruckelshaus*, 742 F.2d 1561, 1565 (D.C. Cir. 1984) (en banc); *Noel v. Chapman*, 508 F.2d 1023, 1030 (2d Cir. 1975).

In fact, for every pitfall I've mentioned, transparency is an essential part of the solution. Simplified choice ensures that consumers understand who is collecting their data and what it is being used for, and that they are given a say in whether that data is collected and how it is used. As we rely on "smart" technology to improve our lives, transparency will enable consumers to know what information is being collected and with whom that information is being shared. As companies move to using big data analytics to make determinations that could adversely affect consumers, transparency will empower consumers to make sure that they are being treated fairly. And expanding consumer access to information that data brokers hold about them will help ensure their practices are transparent. Transparency is the key to accountability, the key to responsible data collection and use, and the key to building consumer trust. Justice Brandeis was surely right when he observed that "[s]unlight is said to be the best of disinfectants."[18]

There is also one risk mitigation technique that is uniquely applicable to the rise of big data — de-identification. De-identification encourages firms that convert "little data" into "big data" to pay attention to stripping out unique identifiers to render the data anonymous. While de-identification isn't foolproof, in our Privacy Report, the FTC has offered an approach to de-identification that seeks to balance the benefits of de-identification with the risks that anonymous data will be re-identified. Our recommendations will be especially important in a big data world.[19]

IV. CONCLUSION

Legend has it that it is a Chinese curse to say to someone "may you live in interesting times." I, for one, am delighted not only be living in an interesting time, but to be heading up an agency that is deeply involved in the policy debates surrounding the innovations that big data will bring about.

Lifeguards have to be mindful not just of the people swimming, surfing, and playing in the sand. They also have to be alert to approaching storms, tidal patterns, and shifts in the ocean's current. With consumer privacy, the FTC is doing just that — we are alert to the risks but confident that those risks can be managed. The FTC recognizes that the effective use of big data has the potential to unleash a new wave of productivity and growth. Like the lifeguard at the beach, though, the FTC will remain vigilant to ensure that while innovation pushes forward, consumer privacy is not engulfed by that wave.

Thank you.

---

[18] LOUIS D. BRANDEIS, OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT 92 (1914).

[19] *See* FTC PRIVACY REPORT, *supra* note 9, at 20-22 (encouraging businesses to (1) use technical and non-technical means to prevent a given data set from being reasonably identifiable, (2) publicly commit not to re-identify the data, and (3) require any downstream users of the data to keep it in de-identified form).