GEORGETOWN LAW

August 15, 2014

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

VIA ELECTRONIC FILING

Re: Comments of Alvaro Bedoya,[1] Center on Privacy & Technology at Georgetown Law, on "Big Data: A Tool for Inclusion or Exclusion? Workshop, Project No. P145406."

This comment focuses on two legal issues surrounding Big Data. First, Big Data's is "big" in large part because companies are increasingly collecting personal data without users' active involvement – or in some cases, consent. Yet while existing privacy law protects user-generated content, it fails to protect these new streams of so-called "passive" data. Perversely, this means that the *less* you know about your data, the *fewer* protections it gets. This is wrong.

Second, industry advocates are promoting the idea that Big Data is essentially too big for consumer control: that consumer controls on the collection of data are impractical and undesirable, and that instead of empowering consumers with a better ability to control the *collection* of their data, the law should instead focus on implementing limits on how that data is *used*.

This strategy is ill-advised. What's more, a movement away from individual controls on collection may be especially harmful for groups that society has at some point deemed undesirable – the poor, the infirm, immigrants, racial and ethnic minorities, and LGBT communities. Privacy is in many ways a shield for the weak. An *exclusive* focus on use limitations would take away that shield and replace it with promises.[2]

---

[1] Executive Director, Center on Privacy & Technology, Georgetown University Law Center. The views expressed here are provided in a personal capacity and do not necessarily reflect those of the Center on Privacy & Technology or of Georgetown Law.

[2] This specific observation, like many others in this second argument, is drawn from a recent comment to the National Telecommunications & Information Administration. *See* Alvaro Bedoya and David Vladeck, *Comments on "Big Data and Consumer Privacy in the Internet Economy,"* NTIA Docket No. 140514424-01, August 5, 2014, at 6 (hereinafter "Bedoya and Vladeck").

600 New Jersey Avenue, NW Washington, DC 20001-2075
PHONE 202-662-9000 FAX 202-662-9444

I. The less you know about your data, the fewer protections it receives.

In May, the White House Big Data report explained that "the volume of information that people create themselves – the full range of communications from voice calls, emails and texts to uploaded pictures, video, and music – pales in comparison to the amount of digital information created *about them* each day."[3] It's true: passively collected data dwarfs data actively generated by users. But a focus on the relative *size* of passively collected data hides its more troubling features. The figure below may explain.

Figure 1. Comparison of Actively and Passively Collected Data.

Kind of Data	Data Actively Created by Users	Data Passively Created *About* Users
Geolocation	GPS data from use of a mapping app or a "check-in"	Cell tower, Wi-Fi and/or GPS data collected by operating system providers in the background (e.g. Apple, Google) Cell tower data automatically collected by wireless carriers from phone calls Cell tower data automatically collected by wireless carriers from Internet use Cell tower, Wi-Fi and/or GPS data collected by a non-location-oriented app
Photo	Photo image	Faceprint derived from image
Email	Email content	Email metadata

Geolocation confirms the Big Data report: passively generated geolocation is more plentiful than actively generated geolocation. A user may check herself into a location or use a mapping app three or four times a day. Unless location services are disabled, Android devices and iPhones automatically collect location data throughout the day and transmit it back to Google and Apple, respectively. One Android phone examined by the *Wall Street Journal* collected Wi-Fi location

---

[3] Executive Office of the President, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES (2014)(hereinafter "Big Data Report") at 2 (emphasis added).

information every few seconds and transmitted it back to Google several times an hour.[4]

But the chart also highlights two problems. First, companies are often generating (or retaining) passive data entirely without a user's knowledge. Users are aware that their location information is being generated when they use Google Maps. But will they remember that Google collects their location information even when they do not use location-aware apps?[5] Similarly, Facebook users know when their friends post photos of them on Facebook. Do they know that Facebook has used those photos to enroll them in a facial recognition database?[6]

Second, in general, the law doesn't protect passively generated data in the same way it protects the actively generated data. This is a result of outdated distinctions between communications *contents* (which get high protections) and *non-contents* (which get low protections) in the Stored Communications Act (SCA).

Under the SCA, Internet companies cannot share the contents of communications without user consent – including, of course, emails. But they have the explicit right to sell or share non-content customer records with "any person other than a government entity" – i.e. any company.[7] Likewise, a photo image clearly constitutes the "contents" of a communication; it could only be shared with consumer consent. It's a little harder to say that for faceprints, meaning that they could potentially be shared freely without user consent.[8]

---

[4] See Julia Angwin and Jennifer Valentino-DeVries, *Apple, Google Collect User Data*, WALL STREET JOURNAL, April 22, 2011; Letter from Bruce Sewell, General Counsel, Apple, to Rep. Ed Markey (D.-Mass.) and Rep. Joe Barton (R.-Tex.) re Apple Inc.'s Response to Request for Information Regarding Its Privacy Policy and Location-Based Services, July 12, 2010.

[5] Both Google and Apple notify users of this collection and take steps to anonymize the data. *See* Android "Location access" Screenshot, Android "Use Google location" Screenshot (presented during device setup), iOS 7 "Location Services & Privacy" Screenshot. But currently, the notifications are less salient than those provided for *apps* accessing location data, which are presented in a detailed app installation screen (Google) or a just-in-time notification (Apple). *See* iOS 7 App Location Permission Screenshot, Android App Installation Screenshot. It appears that Apple's new iOS 8 operating system will provide more detailed location notifications. See *Apple refines location privacy in iOS 8 with new 'While Using' option*, APPLEINSIDER, June 5, 2014.

[6] *See, e.g.*, Letter from Sen. Franken (D.-Minn.) to National Telecommunications & Information Administration, April 2, 2012 at 12-14 (explaining Facebook's creation of a facial recognition database for its Tag Suggestions feature).

[7] 18 U.S.C. § 2702(b)(3),(c)(6).

[8] 18 U.S.C. § 2510(8) defines the "contents" of a communication as "any information concerning the substance, purport, or meaning of that communication," while § 2510(10) defines "communications" to include "any transfer of signs, signals, writing, images, sounds, data, or intelligence of any nature." An image sent via email or posted to a social networking site clearly constitutes the contents of a communication. It is a little harder to see how faceprints – literally metadata (i.e. data about data) measuring the facial characteristics of individuals in a photo – would concern "the

Finally, when you affirmatively open an app to share or determine your location, your location information is arguably the core "content" of your communication. If so, it couldn't be shared without your permission.[9] But in general, location information is considered non-content customer records under the Stored Communications Act, and can thus be shared freely with any non-governmental entity.[10] The only passively generated location data afforded any meaningful protection is that held by wireless carriers as a result of *telephone* calls. The Telecommunications Act prohibits carriers from sharing that data without your consent.[11] But the moment you use that same smartphone to surf the Internet, your wireless carrier can likely share your location data with any other company.[12]

Bountiful, passively generated information is a defining feature of Big Data. But that data is also likely to have been generated without consumers' involvement or even consent – and likely lacks meaningful legal protection. In the world of Big Data, the *less* you know about your data, the *fewer* privacy protections it receives.[13]

This lack of protection for non-content data is especially problematic for health and fitness apps and "wearables" like Fitbit, Nike FuelBand, Jawbone UP, and Samsung Gear Fit. Once activated, these apps and devices are designed to automatically track user activity, be it sleep patterns, a morning jog, or a user's heart rate. Unlike many of the examples of passively generated data cited above, this data is gathered *with* user awareness and consent. But unless it's the contents

---

substance, purport, or meaning" of a communication. In fairness, I am unaware of a court that has addressed this question.

[9] *See* Justin Brookman, *Statement Before the Senate Judiciary Subcommittee on Privacy, Technology, and the Law*, Hearing on "Protecting Mobile Privacy: Your Smartphones, Tablets, Cell Phones, and Your Privacy," May 10, 2011 (hereinafter "Brookman"), at 6 (explaining the potential protection of affirmatively provided or requested location data as "contents" under SCA).

[10] *See ibid. See also* Application of United States for an Order Directing a Provider of Electronic Communication Service to Disclose Records to Government, 620 F.3d 304, 307-308 (3rd Cir. 2010) ("[t]here is no dispute that historical CSLI is a 'record or other information pertaining to a subscriber…or customer,' and therefore falls within the scope of § 2703(c)(1).").

[11] *See* 47 U.S.C. § 222(f); 47 C.F.R. § 64.2001, *et seq*. (extending CPNI rules to cover IP-enabled VoIP services).

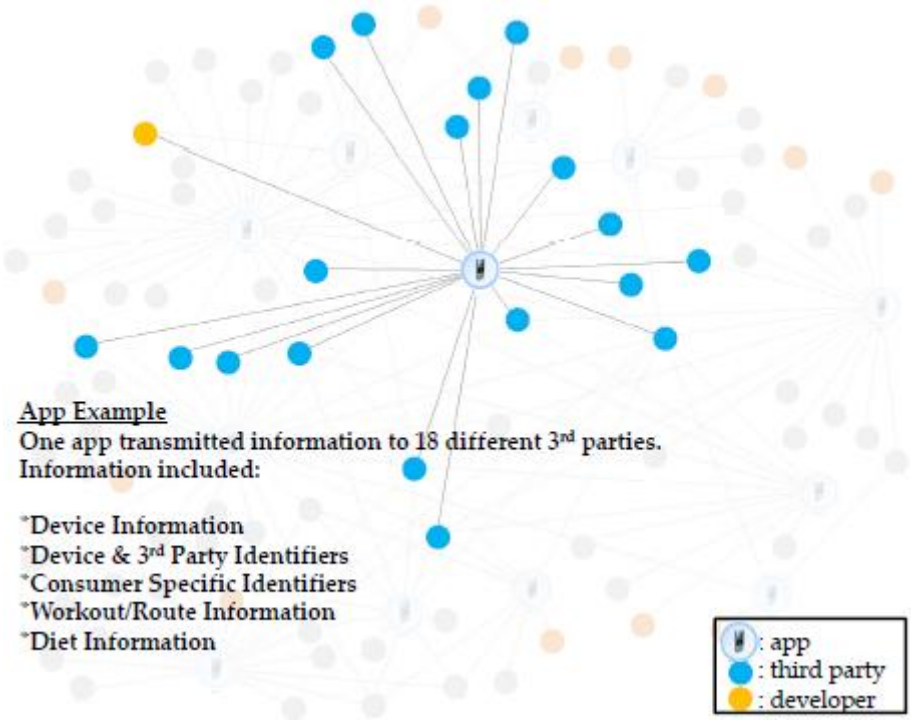
[12] *Appropriate Regulatory Treatment for Broadband Access to the Internet Over Wireless Networks*, Declaratory Ruling, WT Docket No. 07-53, FCC 07-30 (Mar. 23, 2007) (Concurring Statement of Commissioner Michael Copps) at 27, ¶ 3 (carriers offering Title I [Internet] services "appear[] to be entirely free, under our present rules, to sell off aspects of the customer[s'] call or location information to the highest bidder.").

[13] *See* Brookman at 6 (describing "the perverse result that a consumer's information is afforded greater protections when she affirmatively shares sensitive data, as opposed to when her data is shared without her knowledge or consent").

of a communication, the data can likely be shared with any third party company. Are your heartbeats "the contents of a communication"?[14]

App companies take advantage of this lax regulatory scheme. A 2013 audit of 43 health and fitness apps conducted by the Privacy Rights Clearinghouse found that about a third of the apps sent data to third parties who were *not* disclosed in-app or in a privacy policy.[15] A Commission audit of 12 health and fitness apps, including two apps associated with wearable devices, found that those dozen apps disclosed detailed user information – often including users' names, genders, emails, device identifiers, health and workout information – to 76 different third parties.

Figure 2. Commission Analysis of a Health & Fitness App.



One app – pictured in Figure 2 above – sent information to 18 different third parties. One *third party* – an advertising services company – received detailed user data from *four of the twelve* apps that the Commission happened to study. The information included gender, workout data, and keywords like "ovulation," "fertilization," "pregnancy," and "baby." All of this data appears to have been tagged with certain identifiers that would make that data traceable to the same

[14] 18 U.S.C. § 2702(b)(prohibiting the disclosure of the "contents of a communication" absent consumer consent or other criteria).

[15] Linda Ackerman, MOBILE HEALTH AND FITNESS APPLICATIONS AND INFORMATION PRIVACY, REPORT TO CALIFORNIA CONSUMER PROTECTION FOUNDATION, *Privacy Rights Clearinghouse*, July 15, 2013, at 5.

user.[16] This means that one ad company could track a specific user across four apps – *knowing it was tracking that one user across the different apps*.

The lack of protection for non-content data can also create end-runs around other, more powerful privacy laws. This is particularly true for geolocation. Health privacy laws generally prevent an OB/GYN from disclosing the names of his or her patients. But health privacy laws don't cover apps unaffiliated with medical service providers.[17] Investigations by the *New York Times*, the Commission, and the Senate Commerce Committee have found an acute interest in maternity information.[18] "Pregnant women and new parents, after all, are the holy grail of retail," writes the *Times*' Charles Duhigg.[19] It doesn't take much imagination to envision companies who infer pregnancies by geo-tracking users who repeatedly visit an OB/GYN branch. This tracking is already done through other methods.

In other settings, the Commission has recommended that Congress impose collection and sharing protections on certain sensitive data.[20] It has not limited that sensitive information to the *contents* of communications: for example, it has called for requirements that companies obtain users' express affirmative consent before collecting their geolocation data.[21]

I agree with the Commission. Consumers must be given greater control over all data collected about them – regardless of whether that data is "content," and regardless of the passive or automatic nature of that collection. The Commission

---

[16] *See* Transcript, *Consumer Generated and Controlled Health Data*, Federal Trade Commission, Spring Privacy Series, May 7, 2014, at 25-27; Presentation Slides, *Consumer Generated and Controlled Health Data*, Federal Trade Commission, Spring Privacy Series, May 7, 2014 at 27-35.

[17] *See generally*, Department of Health and Human Services, *Summary of HIPAA Privacy Rule*, HHS.gov, accessed on August 15, 2014.

[18] Charles Duhigg, *How Companies Learn Your Secrets*, NEW YORK TIMES, Feb. 16, 2012 (describing Target's use of data mining to predict pregnancies among its shoppers)(hereinafter "Duhigg"); Federal Trade Commission, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY (2014) at 24 (listing data elements gathered by data brokers, including "Expectant or New Parent)(hereinafter "FTC Data Broker Report"); Senate Committee on Commerce, Science, and Transportation, A REVIEW OF THE DATA BROKER INDUSTRY: COLLECTION, USE, AND SALE OF CONSUMER DATA FOR MARKETING PURPOSES, MAJORITY STAFF REPORT, Dec.18, 2014, at 14 ("Equifax maintains approximately 75,000 individual data elements for its use in creating marketing products, including… OB/GYN doctor visits in the last 12 months")(hereinafter "Senate Commerce Report").

[19] Charles Duhigg, THE POWER OF HABIT (2014) at 184.

[20] See FTC Data Broker Report at 52.

[21] *See* Jessica Rich, Director, Bureau of Consumer Protection, Federal Trade Commission, *Statement Before the Senate Judiciary Subcommittee on Privacy, Technology, and the Law*, Hearing on the Location Privacy Protection Act of 2014, June 4, 2014 at 12-13 (supporting legislative measure requiring affirmative express consent before collecting or disclosing geolocation information).

should continue to support congressional efforts to create legal protections for that data. The Commission should also redouble its own enforcement efforts to protect users against the non-consensual collection of their sensitive data.

II. Consumer controls are critical for vulnerable populations.

There has been a concerted push to refocus privacy protections on *use* limitations, rather than controls empowering individuals to prevent collection in the first place. In February 2013, the World Economic Forum and the Boston Consulting Group released a report on Big Data. Based on a year of meetings culminating in a January session in Davos, the report set out a series of "new perspectives" on the use of personal data and highlighted "the need for a new approach" towards that data. Tellingly titled *Unlocking the Value of Personal Data: From Collection to Usage*, the report explained that "[t]he traditional data protection approach… was that the individual is involved in consenting to data use at the time of collection," and argued that this approach was "no longer fit for the purposes for which [it was] designed."[22]

This position is not limited to industry. This May, in paired reports from the Executive Office of the President and the President's Council of Advisors on Science and Technology (PCAST), senior White House officials indicated their own support for this new focus.[23] The Big Data Report actually suggested that it might be *impossible* to give consumers control over the data being collected about them. "[A] sea of ubiquitous sensors, each of which has legitimate uses, make the notion of limiting information collection challenging, if not impossible."[24]

I have argued with David Vladeck that this approach runs opposite to the White House's own Consumer Privacy Bill of Rights, which supported robust – though flexible – individual controls on collection. We also explained that this belief in the inevitability of ubiquitous data collection is, frankly, inaccurate. Data collection sometimes happens by accident, but more often than not it's the result of careful and expensive policy and engineering decisions. Some companies choose to collect data ubiquitously and without consumer consent. Others do not.[25]

---

[22] World Economic Forum, UNLOCKING THE VALUE OF PERSONAL DATA: FROM COLLECTION TO USAGE, Feb. 2013, at 11.

[23] Executive Office of the President, President's Council of Advisors on Science and Technology, BIG DATA AND PRIVACY: A TECHNOLOGICAL PERSPECTIVE (2014) at 49-50 ("Policy attention should focus more on the actual uses of big data and less on its collection and analysis. By actual uses, we mean the specific events where something happens that can cause an adverse consequences or harm to an individual or class of individuals.") (hereinafter "PCAST Report"); Big Data Report at 54.

[24] *Ibid.*

[25] *See* Bedoya and Vladeck at 1-4.

Here, I want to focus on a separate argument we made: that an exclusive focus on use limitations would likely hurt vulnerable communities.

Use limitations work well where the interests of companies and consumers align. For use limitations to be effective, they require that companies and consumers *agree* on which of those potentially harmful uses should be banned. There are a number of areas in which that is the case. But there are also many areas in which consumers' interests diverge significantly from those of the companies that serve them.

There is a growing body evidence that interests may diverge the most for the traditionally disadvantaged.[26] What is harder to recognize, however, is that uses of data that seem acceptable to us today may be found entirely unacceptable later: harmful uses are often deemed harmful only after the fact. Society is especially slow to condemn – or even acknowledge – uses of data that hurt marginalized communities. For example:

- In 1942, Congress repealed the confidentiality protections of the Census,[27] letting the Census Bureau send block-by-block data on the locations of Japanese-Americans to the War Department. Many of them were subsequently rounded up and detained in internment camps.[28]

- After World War II, the U.S. military engaged in wiretapping and mail surveillance to identify and dishonorably discharge gay servicemembers.[29]

- In 1987, the U.S. Public Health Service instituted new mandatory AIDS tests for immigrants, subjecting 500,000 green card applicants to the blood tests annually and barring HIV-positive immigrants from permanent residence.[30]

---

[26] *See, e.g.*, Latanya Sweeney, *Discrimination in Online Ad Delivery*, Jan. 28, 2013 (study showing that a search engine treats first names associated with whites and blacks differently); Jennifer Valentino-Devries, Jeremy Singer, Ashkan Soltani, *Websites Vary Prices, Deals Based on Users' Information*, WALL STREET JOURNAL, Dec. 14, 2012 (investigation revealing that online retailers may quote higher prices to lower-income communities).

[27] Second War Powers Act, PUB. L. NO. 77-507, § 1402, 56 Stat. 186 (1942) (repealed). The measure passed the House on a near-unanimous voice vote. C.P. Trussell, *Wider War Powers Win Vote of House*, NEW YORK TIMES, March 1, 1942.

[28] *See* Steven A. Holmes, *Report Says Census Bureau Helped Relocate Japanese*, NEW YORK TIMES, March 17, 2000 (hereinafter "Holmes"); J.R. Minkel, *Confirmed: The U.S. Census Bureau Gave Up Names of Japanese Americans in WWII*, SCIENTIFIC AMERICAN, March 30, 2007.

[29] *See, e.g.*, Randy Shilts, CONDUCT UNBECOMING: GAYS AND LESBIANS IN THE U.S. MILITARY (2014) at 304.

[30] Bernard Weintraub, *Health Officials Seek AIDS Tests for Immigrants*, May 16, 1987, NEW YORK TIMES.

Most of us now think that these cases of data use or collection were inappropriate, even repugnant. Yet the Census' role in the internment of Japanese Americans was only uncovered in the year 2000, after repeated denials by the Census Bureau.[31] The ban on gays in the military and the HIV travel ban were repealed *only in the last 5 years*.[32] Far too often, today's invidious discrimination was yesterday's national security or public health measure.

This moral lag persists today. And it isn't limited to government. Data brokers have been closely scrutinized for decades.[33] Yet recent investigations by the Commission, the Senate Commerce Committee, and privacy groups have revealed data broker uses of consumer data that are truly reprehensible.[34] A World Privacy Forum investigation, for example, identified consumer targeting lists titled "Aids and Hiv [*sic*] Infection Sufferers," "Rape Sufferers," "Dementia Sufferers" and "Hispanic Payday Loan Responders."[35]

Under what scenario would a victim of sexual assault find it beneficial to be on a marketing list of "rape sufferers?" Why would a company *acting in good faith* send a marketing offer to the mentally incapacitated?

It's just not realistic to think that Congress, companies, and consumers will agree on a set of use restrictions that they all find satisfactory. The American public may *never* make up its mind about women, gay people, immigrants, minorities, the mentally ill, and the poor – or how they and their data should be treated. Individual controls on data collection take that choice out of the hands of companies and the government, and into the hands of the individual.

---

[31] In 1983, a presidential commission concluded that the decisions surrounding the internment of Japanese-Americans were caused by "race prejudice, war hysteria and a failure of political leadership." REPORT OF THE COMMISSION ON WARTIME RELOCATION AND INTERNMENT OF CIVILIANS, PART 2: RECOMMENDATIONS (1983) at 5. Yet the role of the Census remained a secret until the year 2000 – 58 years after the fact. *See* Holmes at 17.

[32] President Obama ended the HIV ban in 2010, calling it "a decision rooted in fear rather than fact." By that time, the U.S. only one of a dozen countries that barred entry to the HIV-positive. *See* The White House, *Remarks by the President at Signing of the Ryan White HIV/AIDS Treatment Extension Act of 2009*, October 30, 2009 (ending the ban effective January 2010). President Clinton announced "Don't Ask, Don't Tell" in 1993. Yet homosexuality investigations continued long afterwards. *See* Roberto Suro, *Navy Sends Agents into Gay Bars*, WASHINGTON POST, June 17, 2000; Tim Weiner, *Military Discharges of Homosexuals Soar*, NEW YORK TIMES, April 7, 1998.

[33] *See* FTC Data Broker Report at 4 (discussing FTC's various investigations into data brokers since 1997).

[34] *See generally*, *ibid*, Senate Commerce Report.

[35] Pam Dixon, Statement before the Senate Committee on Commerce, Science and Transportation, Hearing on *What Information Do Data Brokers Have on Consumers, and How Do They Use It?*, December 18, 2013, at 9, 12, 13, 17.

To be clear, this is an argument in *favor* of individual controls; it is not an argument *against* use limitations. Use limitations are an essential backstop in a world where it is increasingly difficult for consumers to control the collection of their data. Certain privacy invasions can be stopped only through a strong use limitation regime. For example, strong, *ex ante* use limitations could have stopped Target from identifying pregnant women through their purchases.[36] Target would have of course collected the purchase data – the women would have still bought prenatal vitamins and maternity clothing – but a prohibition on medical profiling could have very much kept that genie in the bottle.

If someone truly wants to prevent the harmful use of his or her data, however, there is no better way to do that than to prevent its collection in the first place. As Chairwoman Ramirez has observed, "[i]nformation that is not collected in the first place can't be misused."[37]

III. Conclusion.

Commissioner Ohlhausen said it well: Big Data is invaluable, but "data, even big data, isn't knowledge or wisdom. It can be misleading. Accurately understanding both the benefits and shortcomings of big data technology is critically important."[38] I urge the Commission to take actions that preserve the benefits of Big Data while protecting against potential privacy invasions and disparate impacts, particularly against the traditionally disadvantaged.

---

[36] *See generally* Duhigg.

[37] Chairwoman Edith Ramirez, Federal Trade Commission, *The Privacy Challenges of Big Data: A View from the Lifeguard's Chair*, Keynote Address, Technology Policy Institute, Aspen Forum, Aug. 19, 2013 at 6.

[38] Commissioner Maureen K. Ohlhausen, Federal Trade Commission, *Comments on "Big Data and Consumer Privacy in the Internet Economy,"* NTIA Docket No. 140514424-01, August 5, 2014,