



KEEPING THE INTERNET
OPEN • INNOVATIVE • FREE

www.cdt.org

1634 I Street, NW
Suite 1100
Washington, DC 20006

DATE 10/31/14
TO Federal Trade Commission
Center for Democracy &
FROM Technology
PAGES _____

October 31, 2014

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

RE: Comments to the Federal Trade Commission on the “Big Data: A Tool for Inclusion or Exclusion?” Workshop, Project No. P145406

*“Too many information handlers seem to measure a man
by the number of bits of storage capacity his dossier will occupy.”*
Arthur R. Miller, The Assault on Privacy: Computers, Data Banks, and Dossiers

The Center for Democracy & Technology (CDT) is pleased to submit comments in response to the Federal Trade Commission’s Request For Information (RFI) on the September 15th workshop entitled “Big Data: A Tool for Inclusion or Exclusion?”

At its core, big data is simply information about people that can be rapidly processed by powerful new analytic tools. The potential for big data to inform discussions on critical global issues like public health, climate change, and financial markets is exciting. But the primary application of big data, in a commercial context for marketing purposes, offers neither a solution nor a salve to the world’s problems – it is but one of many technical tools that capitalize on advances in computer speed and improved memory capacity to further business interests. For big data to function, it must have a constant stream of personal information about consumers, often information that includes sensitive matters such as a person’s health, location, and financial status.

Personal privacy, in contrast, is a deeply held belief and human right that is foundational to the concept of individual liberty, autonomy and

MEMO



freedom, one that has been validated by hundreds of years of philosophy and case law. The idea that a loss of control over one's personal privacy is a reasonable price tag for the potential of big data is a fallacy that seems only to give industries an excuse not to ask consumers for permission to use their information. Instead, it's clear that a fair trade in this context would naturally give consumers more leverage over companies who seek to continually feed their data machines. Rather than make a value proposition to consumers for their personal information, many in the business world have moved to a system of hidden pervasive surveillance and collection. Heightened data collection, processing, and retention capacity should indicate that consumers need stronger controls and protections, not weaker and less transparent.

The Fair Information Practice Principles (FIPPs) were originally articulated to address the first wave of big data driven by machine processing and credit scoring in the 1960s and 70s. These principles have guided policymakers in safeguarding privacy for forty years; we believe they are just as relevant today as they were when they were first articulated. We strongly urge the Commission to stress the need for the full range of FIPPs protections over personal data. While big data and algorithmic processing raises interesting and important questions about fairness, harmful uses, and accountability, ultimately consumers may not want to rely on the black box protections offered by companies, and deserve autonomy and agency over the collection and use of their personal information.

Transparency

For all the societal benefits that can come with the availability of large data sets, big data has also given institutions a much greater capacity to monitor, process, and retain data about consumers, all with little to no transparency available to the data contributors themselves. Today, most people have no idea what exists about them in company databases, on digital dossiers collected by data brokers, or the data points that go into complex algorithms used to make determinations about the prices, terms, products, and services they are offered.

We believe that some traditional privacy-enhancing tools are still valuable in pursuing transparency. While few consumers are likely to read even improved privacy policies, they play an important role in public accountability for data processing activities. In theory, these policies make actionable information available to regulators and consumer advocates to assess actual behaviors. Too often, though, these policies do not specifically describe actual practices, and instead reserve broad, vague rights over the collection and use of personal information.

Privacy transparency needs to be improved. Companies should be required to make available within privacy policies specific information about what they're doing with personal information. Moreover, companies should provide contextual, just-in-time notices and other effective consumer notification for data collection or usage practices that would likely surprise users. Without adequate notice and consent provisions, customers who don't approve of what a particular business does won't be able to "vote with their feet" and choose another business with different practices.

For this reason, we strongly support the FTC's use of "material omission" to pursue companies that fail to disclose certain sensitive data practices or gain consent for their collection, such as in the case against Goldenshores Technologies Brightest Flashlight App which deceptively accessed information such as location without users' knowledge. The FTC should continue to aggressively pursue cases like this to promote transparency of practices and to unearth new categories and combinations of information that may require special disclosure and consent, whether in a stand-alone privacy policy or, in some cases, in prominent notices during consent flows.

The FTC should also explore ways that companies could introduce transparency into algorithmic data processing as well. Investigating or holding accountable practices that might lead to discriminatory or disparate impact – such as price discrimination – is made extremely difficult without algorithmic transparency. Achieving transparency in this space has proved challenging. Most consumers also don't have the time, interest, or ability to technically parse their online data trail, and companies have little interest in exposing their proprietary algorithms or the underlying data. To move the discussion forward, we encourage the FTC to consider new ways of thinking about transparency.

The technical details of how a broker transforms "white, 28, female, bicycling, craft beer" into "urban hipster" are largely irrelevant to the regulatory question of whether or not the mechanisms behind the profile are perpetuating discriminatory practices. The FTC should establish conventions around acceptable and unacceptable data streams and data linkages in order to build a consistent regulatory framework that doesn't compromise business models or impede competition. For example, the FTC might consider whether certain streams of data (like Facebook friends and employment history) should be coupled at all.

This guidance would largely echo the core Fair Information Practice Principles. Author Paul Schwartz, for example, recommends that companies using big data analytics "develop reasonable mitigation processes and reasonable remedies as appropriate when analytics lead to decisions that harm individuals,"¹ a suggestion closely aligned

¹ See Paul Schwartz, Data Protection Law and the Ethical Use of Analytics, Privacy & Security Law (Jan. 10, 2011), and Paul Schwartz, Property, Privacy, and Personal Data, 117 HARV. L. REV. 2055, 2096 (2004).

with both the redress element of the individual participation and accountability FIPPs. Recognizing the data quality FIPP, Schwartz also recommends that companies “engage in decision-making based on analytic output that is reasonably accurate” and “only use information that is predictive.”

Potential discriminatory impact assessments are also a promising transparency idea in order to prevent misuse of big data analytics; however, structuring such assessments and implementing oversight programs would require significant time and investment from the private sector. Moreover, without the voluntary release of such assessments and clarity regarding how companies audit their programs, it may difficult to ensure that they are being properly implemented and performed.

Laws such as the Fair Credit Reporting Act (FCRA) and the Equal Credit Opportunity Act (ECOA) offer more useful transparency tools that should be used aggressively by the FTC to enforce broader access and correction rights for consumers. The transparency and correction provisions in the FCRA, for example, provide individuals with a right to obtain a free copy of their consumer report once a year as well as the right to dispute or correct anything in the report, after which the FCRA has 30 days to do a “reasonable” investigation. It makes sense to apply these protections to the gigantic data sets on individuals that are mathematically bound to be incorrect or inaccurate some of the time. Like classifications, errors matter, particularly when they affect vulnerable populations and their ability to obtain credit with reasonable rates. Like traditional Credit Reporting Agencies (CRAs), most people have little to no contact with big data firms and have no remedy for correction or accountability.

Purpose specification and use limitations

Purpose specification and use limitation are two closely related principles that are vital to protecting individual privacy. With respect to consumers, the Administration’s Privacy Bill of Rights well describes the two principles when it says, “Consumers have a right to expect that companies will collect, use, and disclose personal data in ways that are consistent with the context in which consumers provide the data.” Even in the era of big data, purpose specification remains a crucial first step in any system design, requiring entities to detail on what grounds they will collect data and the uses that they plan for it. The use limitation principle requires entities to follow through on the delineated uses and refrain from using the collected data for undisclosed purposes.

Certainly, it will not be possible for companies to spell out every single processing activity — or potential secondary use — even in a standalone privacy policy. However, companies should still endeavor to provide detailed information about the categories of ways it uses data and to provide representative examples. New applications of data that had not been previously disclosed may well be reasonable, so long as they are consistent with the purposes for which the information was taken in the first place. These applications will typically be first-party uses (or done through a dedicated service provider with no independent right to use the data), and should be of a nature such that an ordinary consumer would not find the uses surprising or objectionable. Respect for context must include the understanding that a big data scientist's notion of context may vary wildly from an ordinary consumer's viewpoint and nominal expectations.

While substantive limitations on data usage should play a role in any privacy protection framework, *overreliance* on reasonable commercial use requirements at the expense of individual autonomy and control would in most cases weaken — not strengthen — personal privacy protection.

Relying on responsible use will always be insufficient from a consumer's perspective because it addresses mostly internal accountability and liability, offering no meaningful way for consumers to evaluate practices in the context of their own lives. However, we agree that there should still be some **specific constraints on uses** that are clearly harmful, particularly to vulnerable populations. Use limitations are an important way to make sure entities to follow through on the delineated uses and refrain from using the collected data for undisclosed purposes. Companies must confine their uses of data to the purposes disclosed to consumers and if the company plans to share data collected with a third party, that sharing should be disclosed to consumers in advance, as should the third party's uses (e.g. analytics). Especially in the big data context, entities collecting personal information could very well develop new uses of data in future years that are loosely (if at all) related to the uses that the data was originally collected for. If that happens, entities must at the very least provide transparency about those new uses before they begin. Entities holding data should consider whether the new uses can be performed with de-identified data. They should carefully weigh the potential adverse consequences that may befall individuals from the use of such data and design their programs to avoid such consequences or ensure that they are reliable and justified. However, if data custodians conclude that the new uses can only be performed with identifiable data, and are not contextually related to the purposes for which the data was originally collected, they must seek new consent for those new uses. User expectations — and the potential for user *surprise* — are important indicia for whether a new purpose is contextually related to an older one.

Fundamentally, consumers have the right to control how and when they share information about themselves. This is especially true for consumers in disadvantaged communities, for whom being targeted and tracked can have adverse financial consequences. The United States has a long history of discrimination based on racial and economic profiling and fears of “digital redlining” are far from hypothetical. A report from policy researchers Robinson + Yu² details how even seemingly neutral data can reflect biases in favor of one group over another. In one example cited by the report, a car insurer decided to raise premiums on late-night drivers after remote tracking devices installed by the company determined that drivers on the road past a certain hour in the evening are more often impaired and thus a higher risk to insure. While the decision to bump up rates on these drivers might appear reasonable on the surface, a closer look shows that the greatest financial impact of the rate increase would land squarely on the shoulders of low-income persons of color, a population more likely to work night shifts and live further away from their jobs. The data has effectively created “disparate impact” by penalizing a group based on their socioeconomic status and location. In a report issued for the FTC’s workshop, Peter Swire posits that the ECOA, Fair Housing Act (FHA), and Title VII of the Civil Rights Act of 1964 cover both online and offline marketing activities³. Additionally, Swire suggests that these laws already apply disparate impact analysis to targeted marketing. He also brings up the concept of “steering,” defined by the FHA as “deliberately guiding loan applicants or potential purchasers toward or away from certain types of loans or geographic areas because of race.” In one example, Wells Fargo was found by the Justice Department to have used steering by several methods including its marketing activities and profiling mechanisms to target its subprime lending practices to poor communities and ultimately settled with the department for \$175 million. The company developed software to generate marketing materials for its loan officers that offered a menu of language options. These employees routinely choose a language category entitled “African-American” in order to market subprime loans to poor neighborhoods⁴. Subsequent investigations done by the Justice Department show that it was tactics like this that enabled Wells Fargo brokers to systematically steer thousands of minority borrowers towards higher rates and fees, as well as costlier subprime loans, than white borrowers with similar credit profiles.

² Rieke, Aaron; Robinson, David; Yu, Harlan. *Civil Rights, Big Data, and Our Algorithmic Future* (2014). http://bigdata.fairness.io/wp-content/uploads/2014/09/Civil_Rights_Big_Data_and_Our_Algorithmic-Future_2014-09-12.pdf

³ Swire, Peter. Lessons from Fair Lending Law for Fair Marketing and Big Data (Sept. 2014). http://www.futureofprivacy.org/wp-content/uploads/FairMarketingLessons_WhitePaperFTC.pdf

⁴ Deposition of Tony Paschal, *City of Maryland v. Wells Fargo, N.A.* <https://www.nclc.org/images/pdf/unreported/paschal-decl-balt.pdf>

Consumers need a greater ability to control the information that flows to third-party data brokers that fall outside the Fair Credit Reporting Act. In many if not most cases, sale of personal information to a data broker falls outside the realm of reasonable expectations, and is not consistent with the purpose for which the consumer provided the information. Certainly, merchants today do not do a very good job of explaining secondary transfers of personal information to consumers; even when consumers sign up for a loyalty program or register for a warranty, they are not meaningfully told that their information may be sold to third parties. Reports from the Federal Trade Commission⁵ and the Senate Commerce Committee⁶ reveal that these data broker records often contain highly prejudicial characterizations of people based on their ethnicity, economic status, and even English-speaking ability. As illustrated in the Wells Fargo example, designations like these are not benign for vulnerable consumers.

Using these classifications, the data broker industry facilitates the discriminatory targeting that can lead to groups being penalized based on the perceived risk of that population, whether the penalty is costly subprime loans or the denial of a job. Because the collection, categorization, scoring determinations, and the ways in which these scores are derived, are mostly concealed from consumers, they are powerless to refute, amend or correct information in their profiles. Though the consequences of their practices and products closely mirror those of the regulated consumer scoring industry, data brokers remain **largely unregulated and unaccountable** for the economic harm their industry engenders.

To address this, the FTC should consider investigating scoring practices to unearth circumstances when data is being used as a proxy for discrimination. The Commission should also require data brokers to make current scoring determinations and marketing lists available to the public to dispute, delete or alter if they so choose, and work collaboratively with the Consumer Financial Protection Bureau to investigate how the data used by data brokers might contribute to financial and other impactful determinations such as employment eligibility.

Providing a centralized resource for consumers to find the various companies that are selling profiles about them would be greatly beneficial. In addition, the FTC should require companies to make reasonable default determinations about what data will be collected, how data will be used, how long data will be retained, and with whom data will be shared. Reasonable defaults including making the collection of marketing information opt-out as long as there is an easy and clearly visible way to turn the collection off.

⁵ Federal Trade Commission, *Data Brokers: A Call for Transparency and Accountability* (2014), <http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf>.

⁶ United States Senate Committee on Commerce, Science, and Transportation, *A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes* (2013).

Collection limitations and data minimization

One of the key policy debates around Big Data has been around whether collection limitations and data minimization remain relevant, as getting rid of old data (or not collecting data in the first place) could potentially limit serendipitous uses or insights based on that data. We recognize the valuable societal benefits that could come from secondary research into large data sets, and as we note above, we support secondary uses of previously-collected data that are consistent with ordinary consumer understanding and the purposes for which data was originally given.

On the other hand, mere collection and retention of personal information also presents significant costs and potential for abuse. The **mere possession of consumer data** puts the information at risk and the longer a company holds onto to personal data, the higher the risk becomes. These risks include the increasingly common data breach; internal misuses such as the case of an engineer at a prominent company using his access to spy on user accounts of minors or the recent case in which an AT&T employee obtained customer social security numbers and driver's license numbers; uses that change as company policies change and become inimical to a consumer's best interest; and illegitimate government access.⁷ As such, data minimization is closely related to data security. *Collecting data without a clear (and disclosed) purpose in mind, or the failure to purge old data in accordance with reasonable minimization procedures, should be factors in evaluating whether an entity's data security practices were reasonable.*

As part of their security programs, companies, government agencies and other entities should implement specific, publicly-stated retention periods for data, rather than retaining that information indefinitely in all cases (certainly for some cases, like cloud-stored email, users naturally expect personal data to be stored indefinitely). If entities implement minimization procedures and delete unnecessary, outdated, or irrelevant entries, fewer records will be accessible to unauthorized parties if and when a data breach occurs. By removing identifying information and deleting data after it is no longer needed, companies will both protect their customers' security and promote consumer trust.

Control

Consumers should in many (if not most) cases have the ability to make decisions about how their personal information will be collected, used, and retained, with contextual factors including the sensitivity of the data, the circumstance in which it is collected, and the necessity of it to the core functions of a product. In particular, they should be able to restrict access or use of certain sensitive data or provide permission through affirmative consent, with companies retaining the reciprocal

⁷ Brookman, Justin; Hans, G.S., *Why Collection Matters: Surveillance as a De Facto Privacy Harm*, <http://www.futureofprivacy.org/wp-content/uploads/Brookman-Why-Collection-Matters.pdf>.

ability to deny service if they are unwilling to adhere to a user's stated privacy preferences.

As data flows increase in complexity, it's unrealistic and overly burdensome to force consumers to set data collection, use, and retention preferences on a company-by-company basis. Users should have access to technical mechanisms that allow them to set **global controls** that express their own personal privacy policies (an idea noted in the White House Big Data report⁸), such as the FTC-endorsed Do Not Track technical standard. Furthermore, efforts to trick user agents in order to evade privacy settings are inherently deceptive and should be enforced as such by the FTC. We believe that technical controls will become increasingly important in the modern age and the FTC must be very clear to companies that using hidden mechanisms to subvert user settings is a deceptive practice.

As mentioned above, giving consumers control is a trade off. Apple's recent launch of its HealthKit platform offers an example of the intrinsic exchange that occurs between consumers and companies when customer control is commercialized. Apple rightly gives its customers complete control over what sensitive health information gets shared with what apps and prohibits those apps from selling to restricted data brokers. The decision to recognize the sensitivity of health data and provide true control options, however, necessarily limits Apple's revenue stream from data brokers and means that users of HealthKit will pay directly (as opposed to indirectly through the scooping up of their data) for the service. This exchange creates an imbalance in which privacy is made an expensive commodity, afforded only for the richest and credit-enabled. As the creation of paid-for alternatives to data sharing is likely to be the next phase in the tech marketplace, it's worth considering how restrictions on ad revenue might negatively affect consumers.

Accountability, regulation and legislation

As we have seen over the last ten years, purely self-regulatory models for governing data privacy (bounded only by FTC Section 5 enforcement authority) are inadequate. The inability of the industry to create and enforce meaningful privacy self-regulatory schemes is well documented, none quite as illustrative as in the case of online behavioral advertising. In the early 2000s, with the threat of legislation from Congress looming, industry representatives fought hard for self-regulation and began releasing a series of public initiatives. From the original failure of the Network Advertising Initiative in the early 2000s, to current fragile and, confusing opt-outs that are largely unused by the public and do not fundamentally limit data collection, to the AdChoices icon which is poorly understood by consumers, to the failure to adhere

⁸ *Big Data: Seizing Opportunities, Preserving Values* (May 2014).

http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf

to browser Do Not Track signals, industry self-regulation schemes have been **too weak** placeholders for fundamental rights-based legislation.⁹ On the other hand, prescriptive legislative approaches offer neither enough flexibility nor realistic counters to legitimate constraints to Internet innovation.

We believe that the FTC should back a **principles-based legislative model enforced by the Commission, state attorneys general, and a private right of action**. This would provide a framework with flexibility in the fast-changing tech environment. In fact, we believe that the FTC's use of its existing unfairness authority to pursue companies that fail to provide reasonable security, one of the key Fair Information Practice Principles, has been successful. By targeting bad security practices and delineating rules of the road for other companies to follow, the agency has effectively created legal precedent for this FIPP in US law. The FTC should continue on this regulatory path by reviewing all of the FIPPs when evaluating practices. In addition, the FTC should rigorously investigate current data collection and use practices for new fact patterns and do extensive legal research on how existing law might be invoked to stop data practices that (intentionally or unintentionally) double as proxies for discrimination, such as the ECOA. In one example of the unintended consequences of routine practices online, researchers found that the personal profiles of guests and hosts on lodging website Airbnb.com, which include photos, were providing a mechanism for racial discrimination¹⁰. The study found that hosts identified as white received 12% more for their listings than black hosts despite similar offerings. It also found that black hosts paid a larger penalty for a less desirable location than non-black hosts.

CDT also supports as part of privacy legislation allowing the FTC to endorse industry codes of conduct that are compliant with privacy law, as a mechanism to provide guidance and adaptability to the privacy regulation landscape. Giving the FTC the authority to certify certain practices as a **statutory safe harbor** from privacy enforcement, in conjunction with legislation, would incentivize industries to develop flexible codes that reflect real-world practices. However, legislation should be quite clear that the FTC's approval authority, while given broad discretion, is still contingent upon a code addressing all elements of the Fair Information Practice Principles — including data collection limits, data minimization, and individual control. Regulatory and co-regulatory oversight — as well as express commitments from

⁹ Brookman, Justin, Testimony before the Senate Committee on Commerce, Science, and Transportation, "A Status Update on the Development of Voluntary Do-Not-Track Standards," April 24, 2013, <https://cdt.org/files/pdfs/Brookman-DNT-Testimony.pdf>.

¹⁰ Edelman, Benjamin G. and Luca, Michael. Digital Discrimination: The Case of Airbnb.com (Jan. 10, 2014). Harvard Business School NOM Unit Working Paper, No. 14-054. <http://journalistsresource.org/studies/society/internet/possibilities-online-racial-discrimination-research-airbnb#sthash.LQKNGwqT.dpuf>

companies — will in the short term likely be more effective in preventing big data analytics from being used for discriminatory ends.

The FTC should highlight the importance of individual control and autonomy while signaling out practices that run counter to **an open and fair Internet**, such as price discrimination, the collection and use of sensitive categories improperly, and opacity of industry practices and algorithms that prohibit effective accountability and responsible innovation.

Sincerely,

Michelle De Mooy
Deputy Director, Consumer Privacy Project
Center for Democracy and Technology