

2017

Trusting Big Data Research


Neil M. Richards

Washington University in St. Louis School of Law, nrichards@wustl.edu

Woodrow Hartzog

Boston University School of Law

Follow this and additional works at: https://openscholarship.wustl.edu/law_scholarship

 Part of the [Commercial Law Commons](#), [Consumer Protection Law Commons](#), [Internet Law Commons](#), [Legal Studies Commons](#), [Privacy Law Commons](#), and the [Science and Technology Law Commons](#)

Repository Citation

Richards, Neil M. and Hartzog, Woodrow, "Trusting Big Data Research" (2017). *Scholarship@WashULaw*. 542.

https://openscholarship.wustl.edu/law_scholarship/542

This Essay is brought to you for free and open access by the Law School at Washington University Open Scholarship. It has been accepted for inclusion in Scholarship@WashULaw by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

TRUSTING BIG DATA RESEARCH

Neil Richards* & Woodrow Hartzog**

ABSTRACT

Although it might puzzle or even infuriate data scientists, suspicion about big data is understandable. The concept does not seem promising to most people. It seems scary. This is partly because big data research is shrouded in mystery. People are unsure about organizations' motives and methods. What do companies think they know about us? Are they keeping their insights safe from hackers? Are they selling their insights to unscrupulous parties? Most importantly, do organizations use our personal information against us? Big data research will only overcome its suspicious reputation when people can trust it.

Some scholars and commentators have proposed review processes as an answer to big data's credibility problem. It is possible that a review process for big data research could provide the oversight to ensure the ethical use of data we have been hoping for, applying sensible procedural rules to regularize data science. But procedure alone is not enough. In this essay, we argue that to truly protect data subjects, organizations must embrace the notion of trust when they use data about or to affect their human users, employees, or customers. Promoting meaningful trust involves structuring procedures around affirmative, substantive obligations designed to ensure organizations act as proper stewards of the data with which they are entrusted. To overcome the failures of a compliance mentality, companies must vow to be Protective, Discreet, Honest, and above all, Loyal to data subjects. Such commitments backed up by laws will help ensure that companies are as vulnerable to us as we are to them. When we know we can trust those using big data, the concept might not seem so scary after all. We will disclose more and more accurate information in safe, sustainable ways. And we will all be better off.

* Thomas and Karole Green Professor of Law, Washington University School of Law, Affiliate Scholar, The Center for Internet and Society at Stanford Law School, Affiliated Fellow, Yale Information Society Project.

** Starnes Professor of Law, Samford University Cumberland School of Law, Affiliate Scholar, The Center for Internet and Society at Stanford Law School.

I. INTRODUCTION

To those in industry, policymakers, and academia, the scientific tools colloquially referred to as “big data” are exciting. Data science has the potential to change our lives for the better, to be, in the words of two scholars, “a revolution that will transform how we live, work, and think.”¹ Powerful algorithms can be combined with mountains of personal information to produce insights and predictions to help solve virtually any problem we can think of, but for most people, big data does not seem promising. It seems scary.

The public paranoia is understandable because big data research is shrouded in mystery. People are unsure about organizations’ motives and methods. What do companies know about us? Are they keeping their insights safe from hackers? Are they selling their insights to unscrupulous parties? Most importantly, do organizations use our data against us? Like other Information Age concepts that cause privacy anxiety, like biometrics, automated technologies, and the Internet of Things, big data research will only overcome its suspicious and clandestine reputation when people can come to trust it.

Scholars, commentators, companies, and even the White House have proposed review processes of various sorts as an answer to big data’s credibility problem.² These review processes could be implemented by those conducting big data research to provide the same kind of protections ensured by university Institutional Review Boards (IRBs), which seek to make sure that those conducting research on humans minimize risk, follow ethical research principals, and receive informed consent.³ Proposals of this sort are well worth investigating and debating. It is certainly possible that structural review processes for big data research could provide the oversight to ensure the ethical use of data we have been hoping for; however, procedure alone will not be enough. We also need the right substantive safeguards.

Procedural approaches to big data research pose at least three problems. First, organizations risk falling into a compliance mentality, seeking only to satisfy procedural (instead of substantive) demands. When this happens, organizations can cut corners and do the least

1. VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, *BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK* 11 (2013).

2. *See, e.g.*, Consumer Privacy Bill of Rights Act of 2015, S. 1158, 114th Cong. § 103(c) (1st Sess. 2015); Ryan Calo, *Consumer Subject Review Boards: A Thought Experiment*, 66 *STAN. L. REV.* 97 (2013); Jules Polonetsky et al., *Beyond the Common Rule: Ethical Structures for Data Research in Non-Academic Settings*, 13 *J. TELECOMM. & HIGH TECH. L.* 333, 334 (2015); Mike Schroepfer, *Research at Facebook*, *FACEBOOK* (Oct. 2, 2014), <http://newsroom.fb.com/news/2014/10/research-at-facebook/>.

3. *See supra* note 2 and accompanying text.

amount possible, and the spirit of the law becomes subservient to the letter of the law. Second, the dominant procedural approaches are all premised upon the manufacturing or extraction of consent. While consent models might work well at the one-to-one level between human researchers and subjects, they simply do not scale for either institutions or humans. Large organizations can manufacture meaningless consent via boilerplate terms that no one reads nor should be expected to read.

Moreover, meaningful informed consent would be so burdensome at scale as to make most big data research (and indeed most uses of data) cost prohibitive. Just imagine employees at Facebook offering a one-on-one chat session with each of its 1.86 billion monthly active users to answer questions they might have about the research it is conducting with user data.⁴ Similarly, from the perspective of individual humans in information relationships, the typical modern human (whether we call her a “user,” a “consumer,” or a “citizen”) will have scores or even hundreds of relationships with social networks, search engines, cloud service providers, antivirus software manufacturers, internet service providers, hardware and operating system manufacturers of home or mobile computers, tablets or phones, airlines, taxi and transportation companies, and accounts with online, offline, and hybrid merchants and websites offering a bewildering array of services and an equally bewildering array of practices with respect to their personal data. This is before we get into information relationships with which users may be unaware, such as those with data brokers and advertiser networks. From the perspective of either humans or institutions, the model of careful informed consent on which procedural methods rest is impractical at best (and farcical and fraudulent at worst).

Fair Information Practice Principles (FIPPs), including data minimization, control, security, and notice, have played a critical role in our “small data” world⁵ but a few of the principles struggle with big data. The FIPPs that are focused on the procedural requirements of notice and consent do not scale well because they eventually overwhelm people. Giving people control over their data can feel debilitating or like a trap given the difficulties in ascertaining risk and lack of real op-

4. *Company Info*, FACEBOOK, <http://newsroom.fb.com/company-info/> (last visited Feb. 13, 2017) (listing the number of users as of December 31, 2016).

5. See ROBERT GELLMAN, FAIR INFORMATION PRACTICES: A BASIC HISTORY 14–15 (June 17, 2016), <http://bobbegelman.com/rg-docs/rg-FIPShistory.pdf>.

tions. Consent and harm are good but are mismatches as they do not really remedy what we are concerned about.⁶

There is a better model for big data research than rote and meaningless “compliance and consent.” In this essay, we argue that to truly protect data subjects, organizations must embrace the notion of trust as a guiding principle in their processing of personal data. This involves structuring procedures around affirmative obligations designed to ensure organizations act as adequate⁷ stewards of the data with which they are entrusted. Right now, people distrust companies that are trying to leverage personal information with big data techniques. The Target Pregnancy story is the evergreen anecdote justifying big data paranoia,⁸ but consider also the backlash against Facebook’s new Artificial Intelligence assistant—“M”—and its infamous “mood study” that explored the algorithmic possibilities of emotional manipulation.⁹

Procedures to ensure ethical big data research are necessary, but alone they are wholly insufficient. Without the right substantive rules, ethical review processes risk becoming mere formalities on the road to data strip mining. We can do better than creating mere incentives for

6. See Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 20 STAN. TECH. L. REV. 431, 436–37 (2016); see also Neil Richards & Woodrow Hartzog, *Privacy’s Trust Gap*, 126 YALE L.J. 1180 (2017). Other scholars have also recently started to examine the relationship between privacy and trust. See generally Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183 (2016); Kiel Brennan-Marquez, *Fourth Amendment Fiduciaries*, 84 FORDHAM L. REV. 611 (2015); Dennis D. Hirsch, *Privacy, Public Goods, and the Tragedy of the Trust Commons: A Response to Professors Fairfield and Engel*, 65 DUKE L.J. ONLINE 67 (2016); Robert H. Sloan & Richard Warner, *“I’ll See”: How Surveillance Undermines Privacy by Eroding Trust*, 32 SANTA CLARA HIGH TECH. L.J. 221 (2016); Ari Ezra Waldman, *A Breach of Trust: Fighting ‘Revenge Porn,’* 102 IOWA L. REV. 709 (2017); Ari Ezra Waldman, *Manipulating Trust on Facebook*, 29 LOY. CONSUMER L. REV. 175 (2016); Ari Ezra Waldman, *Privacy as Trust: Sharing Personal Information in a Networked World*, 69 U. MIAMI L. REV. 559 (2015); Ari Ezra Waldman, *Privacy, Sharing, and Trust: The Facebook Study*, 67 CASE WEST. RES. L. REV. 193 (2016).

7. We use this word consciously with its echoes of the “adequacy” requirement of European Data Protection Law. See Press Release, Court of Justice of the European Union, The Court of Justice Declares that the Commission’s US Safe Harbour Decision Is Invalid (Oct. 6, 2015), <http://curia.europa.eu/jcms/upload/docs/application/pdf/2015-10/cp150117en.pdf>; see also Robert Levine, *Behind the European Privacy Ruling that’s Confounding Silicon Valley*, N.Y. TIMES (Oct. 9, 2015), <http://nyti.ms/1NsQP6a>; cf. Schrems v. Data Protection Comm’r, ECLI:EU:C:2015:650 (E.C.J. 2015), <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:62014CJ0362&from=EN>.

8. See, e.g., Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES MAG. (Feb. 16, 2012), <https://nyti.ms/2jEboTD> (detailing Target’s strategy of identifying women in their second trimester of pregnancy).

9. See, e.g., Neil Richards & Woodrow Hartzog, *Facebook’s New Digital Assistant ‘M’ Will Need to Earn Your Trust*, GUARDIAN (Sept. 9, 2015), <https://www.theguardian.com/technology/2015/sep/09/what-should-we-demand-of-facebooks-new-digital-assistant>.

compliance with procedures, and we must do better if data science is to achieve its touted potential.¹⁰ To overcome the failures of a compliance mentality, we argue that companies (and other institutions processing personal data) must adhere to four substantive commitments when they process personal data of people with whom they have an information relationship. Such institutions must vow to be Protective, Discreet, Honest, and above all, Loyal to data subjects. Such commitments backed up by laws will help ensure that companies are as vulnerable to us as we are to them. When we know we can trust those using big data, the concept might not seem so scary after all. We will disclose more and better information in a safe, sustainable way, and we will all be better off.

II. PROTECTION

Big data requires distribution and linked research projects. Data must be stored in order to exist, and it must be accessible in order to be useful. At scale, big databases are honeypots that demand robust data security. One key component of data security is the old FIPPs of data minimization.¹¹ Data that does not exist cannot be leaked or hacked. Of course, data that does not exist cannot be exploited, and it is for this reason that data minimization is seen by some as anathema to good data science.¹² Such a “collect it all and let the algorithm sort it out” mindset makes big data riskier than your average dataset. To make matters worse, in the age of the data breach, there are three kinds of organizations: those that know they have been breached, those on the precipice of breach, and those that have been breached discreetly and just do not know it yet.¹³

The FIPPs have long recognized the need for data security. “Data Protection” in its most literal sense implies data security even more than it does privacy. The first “code of fair information practices” issued by the famous 1973 Advisory Report of the Committee of the Department of Health, Education, and Welfare concluded that “[a]ny

10. See Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393, 393 (2014).

11. See GELLMAN, *supra* note 5, at 15.

12. Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 259 (2013) (“The big data business model is antithetical to data minimization.”).

13. This is with apologies to former FBI Director James Comey, who famously quipped on CBS’s *60 Minutes* that “there are two kinds of big companies in the United States. There are those who’ve been hacked by the Chinese and those who don’t know they’ve been hacked by the Chinese.” Scott Pelley, *FBI Director on Threat of ISIS, Cybercrime*, CBS NEWS (Oct. 5, 2014), <http://www.cbsnews.com/news/fbi-director-james-comey-on-threat-of-isis-cybercrime/> (transcribing the CBS television broadcast).

organization creating, maintaining, using, or disseminating records of identifiable personal data must assure the reliability of the data for their intended use and must take reasonable precautions to prevent misuse of the data.”¹⁴ Subsequent versions of the FIPPs have invariably held data security up as a foundational requirement of the ethical management of any personal data processing or storage, especially as the Age of the Data Breach has arrived.¹⁵

There is a problem with the way in which the obligation of “security” has manifested in law and corporate business practices. An obligation of “security” is an ambiguous one, and a company can take different approaches to data security. One of the most common approaches has been compliance-based, which involves checking off a list of reasonable technical and administrative measures, merely to avoid liability. This view manifests itself in demands that data security requirements be bright-line rules that contain safe harbors for practices that meet common security practices in the industry. A compliance mentality dictates that once a certain threshold of security has been met, the problem is solved.

But procedure should be just one part of an organization’s approach to data security. A security requirement based on compliance with “industry standards” is ineffective and insecure if industry standards are themselves insufficient. Such an approach would be akin to resting workplace safety rules on the labor practices of sweatshops at the turn of the last century. Companies certainly deserve clear and fair guidance from the law, but the protection of personal data cannot be merely about avoiding liability. Instead of a procedurally watered-down version of the FIPP of “security,” we suggest instead a robust duty of Protection should apply to warehouses of personal data.

This principle looks not only to traditional safeguards and procedures, but it also attempts to protect information as it flows downstream. Industry standards are a good starting point for this process; however, the true test of adequacy for companies that embrace the protection principle will be whether organizations took all steps within reason to protect a subject’s information in order to preserve trust. Did an organization only protect data until it was transferred to a third party? Did a company follow the pack even when industry standards were insufficient, such as with the widespread failure of the

14. SEC’Y’S ADVISORY COMM. ON AUTOMATED PERS. DATA SYS., U.S. DEP’T. OF HEALTH, EDUC. & WELFARE, RECORDS, COMPUTERS AND THE RIGHTS OF CITIZENS 41, 50 (1973) [hereinafter 1973 FIPPs], <https://epic.org/privacy/hew1973report/>.

15. See GELLMAN, *supra* note 5, at 14–15.

password and lack of adoption of two-factor authentication?¹⁶ Apple's lack of two-factor authentication in light of the notorious iCloud hack is a prime example of a company failing to fully embrace the protection principle.¹⁷

Instead of making security merely a compliance obligation, organizations should focus on creating value by protecting data. Robust efforts at data protection will help gain and keep the trust of data subjects, which will generate more disclosures within a safe, sustainable relationship. Companies that go beyond mere checklists and make data protection a priority through the allocation of resources and time should be considered trustworthy.

Companies can directly benefit from robust security as well. Breached data can be changed and corrupted. Data integrity is a key principal of data security. It also directly implicates the financial interests of companies entrusted with personal information. Without integrity, data is useless. Corrupt data can even be harmful if the corruption goes undiscovered. Thus data protection does more than generate trust, it directly benefits companies by ensuring that the information they are entrusted with remains useful.

III. DISCRETION

One of the major fears people have over big data research is who will have access to insights gleaned from big data. One of the popular big data anecdotes involves the insight that people who buy pads for their furniture are better insurance risks.¹⁸ In addition, credit companies have been looking to Facebook profiles as part of their decision to extend credit.¹⁹ Are they buying big data insights from other companies as well? This is part of the reason why people have trust issues with big data. Put simply, people do not know what happens to their

16. See, e.g., Daniel Solove & Woodrow Hartzog, *Should the FTC Kill the Password? The Case for Better Authentication*, 14 BLOOMBERG BNA PRIVACY & SEC. L. REP. 1353 (2015).

17. Arik Hesseldahl, *Apple Says It Is "Actively Investigating" Celeb Photo Hack*, RECODE (Sept. 1, 2014, 1:49 PM), <http://recode.net/2014/09/01/apple-says-it-is-actively-investigating-celeb-photo-hack/>; Paul Tamburo, *WWDC 15: Apple Ensures "The Fapping" Won't Happen Again with Two-Factor iCloud Authentication*, CRAVE (June 8, 2015), <http://www.craveonline.com/design/864955-wwdc-15-apple-ensures-fapping-wont-happen-two-factor-icloud-authentication#scWUL2DZMg8QT7C8.99>.

18. Jonathan Shaw, *Why "Big Data" Is a Big Deal*, HARVARD MAG., Mar.-Apr. 2014, at 30, 31 ("Credit-card companies have found unusual associations in the course of mining data to evaluate the risk of default: people who buy anti-scoff pads for their furniture, for example, are highly likely to make their payments.").

19. *How Your Facebook Profile Can Affect Your Credit*, ATLANTA J.-CONST. (Nov. 2, 2015, 6:12 PM), <http://www.ajc.com/news/news/national/how-your-facebook-profile-can-affect-your-credit/npD9X/>.

information once it is (as they often colloquially put it) “out there,” which makes them fearful and distrustful of both the system and those who may be using the data.

Like Security, one of the bedrock FIPPs is nondisclosure. The 1973 Advisory Report of the Department of Health, Education and Welfare mandated that “[t]here must be a way for an individual to prevent information about him that was obtained for one purpose from being used or made available for other purposes without his consent” and further placed an obligation on organizations holding personal data to “take precautions to prevent misuse of the data.”²⁰ Similar requirements run through all sincere restatements of the FIPPs, including the recent White House “Privacy Bill of Rights.”²¹ These requirements are part of an ancient tradition in our law, protecting shared personal information through duties of confidentiality, whose overtones of “confidence” resonate directly with notions of “trust.”²²

In practice, however, our longstanding commitments to confidentiality, or at least nondisclosure, have been gutted by the procedural approach to the FIPPs. Notions of confidentiality have been replaced by a compliance-based system of “notice and choice,” in which the default of nondisclosure has increasingly flipped to a default of disclosure unless a person takes affirmative steps to opt out of data sharing. In practice, opting out of data sharing often requires the termination of an information relationship. As we have argued elsewhere in much greater detail, our commitments to nondisclosure have been replaced by an illusion of control of our personal data.²³

In the big data context, these problems threaten to become even worse. Confidentiality regimes are typically premised upon the confidant not disclosing information received by the confider. Big data, however, involves predictions and insights derived from information (some of it possibly confidential). Theoretically, a confidant following a compliance “letter, not spirit, of the law” approach might be able to disclose predictions and insights without technically breaching confidentiality. A trust-based regime would apply to more than just confidential information, extending an obligation to be discreet regarding any personal information, disclosed or surmised. The data subject

20. See 1973 FIPPs, *supra* note 14, at xx–xxi.

21. See Press Release, White House, Fact Sheet: Plan to Protect Privacy in the Internet Age by Adopting a Consumer Privacy Bill of Rights (Feb. 23, 2012), <http://www.whitehouse.gov/the-press-office/2012/02/23/fact-sheet-plan-protect-privacy-internet-age-adopting-consumer-privacy-b>. See generally GELLMAN, *supra* note 5.

22. Neil M. Richards & Daniel J. Solove, *Privacy's Other Path: Recovering the Law of Confidentiality*, 96 GEO. L.J. 123, 135 (2007).

23. Richards & Hartzog, *Taking Trust Seriously in Privacy Law*, *supra* note 6, at 444–47.

would feel betrayed if disclosed indiscreetly or to wrongful persons. The broad scope of this obligation would be mitigated by the fact that a duty of discretion means limited disclosure is allowed to other trustworthy parties and for purposes not adverse to the data subject's interest. The important point is that discretion is more permissive than a fiduciary duty of confidentiality, but it should apply more broadly. An obligation of discretion keeps companies from falling into a potentially cynical compliance mentality that looks for technical loopholes to disclosure obligations. The onus instead is put on the trustee for reasonable disclosure under the circumstances.

IV. HONESTY

Since its inception, data protection law has also placed obligations of notification on holders of personal data. After all, it would be virtually impossible to give data subjects the ability to control how their data is used if they have no idea that the data exists in the first place. The original 1973 FIPPs had two requirements along these lines. They mandated that first, “[t]here must be no personal data record-keeping systems whose very existence is secret,” and second, “[t]here must be a way for an individual to find out what information about him is in a record and how it is used.”²⁴

These requirements might seem substantive, but in practice they have also fallen prey to a compliance-focused proceduralism. The dominant means of modern privacy regulation in the United States is a thin regime of so-called “notice and choice,” in which data practices are described in long privacy policies that all too often manage to be both vague and wordy at the same time.²⁵ The multiple problems with this approach are well-documented; foremost among its flaws is the fact that the onus of finding and reading these policies is placed upon individual users, even though reading the privacy policies we encounter on a daily basis would literally take days to complete.²⁶ The impracticality of notice and choice means that in practice its regime is a

24. See 1973 FIPPs, *supra* note 14, at xx.

25. Florencia Marotta-Wagner, *Does “Notice and Choice” Disclosure Regulation Work?* 26 (NYU L. Sch., Working Paper No. 13, 2015), <https://www.law.umich.edu/centersandprograms/lawandeconomics/workshops/Documents/Paper13.Marotta-Wurgler.Does%20Notice%20and%20Choice%20Disclosure%20Work.pdf>.

26. See Richards & Hartzog, *supra* note 6, at 444–47; Daniel J. Solove, *Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880 (2013); Alex C. Madrigal, *Reading the Privacy Policies You Encounter in a Year Would Take 76 Work Days*, ATLANTIC (Mar. 1, 2012), <http://www.theatlantic.com/technology/archive/2012/03/reading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851/>; see also Aleecia M. McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 I/S: J.L. & POL’Y INFO. SOC’Y 543 (2008).

wholesale abandonment of the purpose of the FIPPs. When applied to the complex science and applications of big data, a notice and choice regime relying on the efforts of users would be even more laughable.

Instead of our current, cynical regime of notice and choice, we believe that institutions dealing with big data should adopt the substantive principle of Honesty. Instead of the constructive, lawyerly, and realistically deceptive practice of “notice,” organizations should take steps to be honest with their users about what they do and why they do it. Such an obligation would transcend the compliance-focused, legalistic notion of notice by focusing on what users actually understand rather than on what institutions can get away with by burying their activities in the fine print. Honesty takes the obligation of understanding away from individual users, and places the obligation on the organization to explain and to be understood. Honesty takes the fiction out of *constructive notice* to require *actual notice*. But unlike “consent” process models, honesty does not serve to transfer the risk of loss onto data subjects. Rather, it serves as a trust-building mechanism to reduce anxiety and encourage mutually-beneficial, reciprocal disclosure.

An obligation of Honesty also serves an additional function of forcing companies to take stock of their information practices in order to be accurate when keeping individuals informed.²⁷ When users feel that their data is being processed in ways they poorly understand, they can feel like they are playing a kind of legalistic whack-a-mole. It should be no surprise that such disempowering tactics foster user suspicion and distrust. In contrast, when institutions explain how and why they are using data, this honesty can serve as an important foundation of trust.

V. LOYALTY

Protection, Discretion, and Honesty are substantive principles with long antecedents in data protection and privacy law around the globe. Even organizations that are Protective, Discreet, and Honest might find that these activities are not enough to build trust. There is the nagging suspicion among many users that organizations are still exposing their data so that the organization will take advantage of them and

27. Peter P. Swire, *The Surprising Virtues of the New Financial Privacy Law*, 86 MINN. L. REV. 1263, 1264–65 (2002) (“[C]ritics have largely overlooked . . . important benefits from these notices. Perhaps most significantly, publication of the notices and the new legal obligation to comply with them have forced financial institutions to engage in considerable self-scrutiny as to their data handling practices.”).

enrich itself at their expense. We see this fear of betrayal in the common adage that Facebook users, for example, are not the customers but instead the product—groomed, curated, and offered up unsuspectingly to Facebook’s fee-paying advertiser clients.²⁸ Philosopher of technology Nicholas Carr puts it more bluntly, calling such practices “digital sharecropping.”²⁹

To remedy this problem, we offer a fourth substantive principle to build trust in information relationships involving big data. We call this principle Loyalty, the idea that organizations should not enrich themselves at the expense of their trusting users. Although Loyalty is a new value in the big data policy debate, it has very old antecedents in the common law. Because it is new, it has not yet had a chance to be watered down by cynical proceduralism. We must not allow that to happen. Loyalty is the linchpin of trust, for without it, users are justified in being suspicious of every use of their data and decision about them a company makes. Without loyalty, users are exposed to self-dealing and betrayal. Trust, as we have argued elsewhere, is the key element to the kind of sustainable information relationships that are the best future for the digital society.³⁰

What does loyalty mean in the practice of big data research? It means that companies must put the interests of their users ahead of the narrow short-term profit motive when deploying the products of big data research. It is ethical to mine the data trove of user information to recommend movies, television shows, or even new services the user may enjoy, but it is not ethical to use that data to discover their reservation price. Consider the trove of data that Amazon holds on behalf of its users. It can use that information to suggest new movies they might want to rent or accessories for the products they already own. These kinds of offerings can be loyal and create value for both the company and its customers. A much harder case would be where Amazon uses the fruits of its data research to find out the highest price each customer would want to pay for each such product. Another unjustified use of such data would be to uncover and exploit compulsive behaviors or addictions, like personalizing shopping experiences for suspected gambling addicts to resemble a slot machine. Further, it would be a betrayal of the duty of Loyalty for Amazon to

28. Olivia Solon, *You Are Facebook’s Product, Not Customer*, WIRED UK (Sept. 21, 2011), <http://www.wired.co.uk/news/archive/2011-09/21/doug-rushkoff-hello-etsy> (attributing the phrase to author Douglas Rushkoff).

29. Nicholas Carr, *Digital Sharecropping*, ROUGH TYPE (Dec. 19, 2006), <http://www.roughtype.com/?p=634>.

30. See generally Richards & Hartzog, *Taking Trust Seriously in Privacy Law*, *supra* note 6.

sell insights about its customers to those who might harm them, whether data brokers, insurance companies, employers, or others.

Importantly, the duty of loyalty would allow for limited self-dealing within reason, so long as it did not come at the expense of the data subject. Thus, companies would still be allowed to make some use of entrusted information. This obligation is a slightly more relaxed, but it is also a more broadly applicable standard to that of fiduciaries. If it is calibrated correctly, a duty of loyalty could prevent a number of information misuses at the heart of big data anxiety. Like the tort of negligence and the generally accepted standard for data security, what is reasonable under the circumstances is entirely dependent upon context, and it must be developed over time. It is clear, however, that without some sense that organizations will remain true to those who trust them with their personal information, big data anxiety and distrust will only continue.

VI. CONCLUSION

Big data paranoia is justified, yet ultimately unproductive. We should look for a way forward that allows us to explore some of the potential benefits of big data, while avoiding some of its obvious risks. We believe that trust should be the lodestar of big data ethics and law, and that in order to promote trust, organizations must commit to Protection, Discretion, Honesty, and Loyalty. To protect against the inevitable bad or short-sighted actors, the law should mandate these protections where necessary. Specifically, the law should shift away from the FIPP based notions of consent that dominate review process regimes like IRBs. Consent at scale only shifts the risk of harm away from companies and makes people skeptical. If big data research is ever going to get off the ground, organizations must give people a reason to trust them.