University of Dayton Law Review

Volume 40 | Number 2 Article 3

10-1-2015

The Unexamined Life in the Era of Big Data: Toward a UDAAP for Data

Sean Brian
University of Utah

THE UNEXAMINED LIFE IN THE ERA OF BIG DATA: TOWARD A UDAAP FOR DATA

Sean Brian[1]

"*[A]n unexamined life is not worth living*."[2]

I. INTRODUCTION

With few exceptions, a user can sign up for an account and receive content and services without providing anything more than the credentials they wish to use at the site.[3] Yet the companies that provide these services

[1] Sean Brian, J.D. Candidate 2015, S.J. Quinney College of Law, University of Utah.
[2] PLATO, THE APOLOGY OF SOCRATES 77 (D. F. Nevill trans., 1901).
[3] Ethan Zuckerman, *The Internet's Original Sin*, THE ATLANTIC (Aug. 14, 2014, 6:04 AM), http://www.theatlantic.com/technology/archive/2014/08/advertising-is-the-internets-original-sin/376041/.

are valued in the billions of dollars.[4] Users have a vague sense that Facebook, Google, and other Internet companies make money through advertising, and indeed the sheer number of users is a big factor in attracting advertisers, but this is only a small part of the picture.[5] The other, and perhaps even more valuable, source of revenue is the data that these companies collect about users' preferences and demographics.[6] In sum, data has become the currency of the Internet.

An entire industry has developed around purchasing, analyzing, and even combining the data collected by Internet companies.[7] This industry is largely responsible for allowing the Internet to evolve from a means of sharing information between universities into what it has become today.[8] Both the gathering and use of data collected online have begun to raise privacy concerns with regulators.[9] They question whether consumers are really giving informed consent as they sign up for these services, and what limits should be placed on the use and sale of consumer data, including whether consumers should have the opportunity to opt out.[10]

Overly burdensome regulation could end the era of free online services as well as limit access to the wealth of data they produce.[11] Currently, a patchwork of privacy regulations and laws have left marketing associations trying to fill the void with industry-best-practices to provide clear guidelines on gathering and using data while respecting consumers'

---

[4] *See Market Value of the Largest Internet Companies Worldwide as of May 2014 (in Billion U.S. Dollars)*, STATISTA, http://www.statista.com/statistics/277483/market-value-of-the-largest-internet-companies-worldwide/ (last visited Mar. 24, 2015).

[5] Mark Sullivan, *How Will Facebook Make Money?*, PCWORLD (Jun. 14, 2010, 10:00 PM), http://www.pcworld.com/article/198815/ssss.html.

[6] *Id.*; *see also* Thiag Loganathan, *Are You Using Your Data to its Full Potential?*, DMI, http://dminc.com/blog/using-data-fullest-potential/ (last visited Feb. 14, 2015). *See generally* DATAFLOQ, https://datafloq.com/ (last visited Mar. 24, 2015).

[7] *See* Loganathan, *supra* note 6; Jennifer L. Rathburn & Simone Colgan Dunlap, *Big Data, Big Risk: Strategies to Mitigate Risks Associated with Data Monetization*, INSIDE COUNSEL (Dec. 18, 2014), *available at* http://www.insidecounsel.com/2014/12/18/big-data-big-risk-strategies-to-mitigate-risks-ass.

[8] Zuckerman, *supra* note 3; Erin Griffith, *An Audacious Plan to Fix the Internet's Original Sin*, FORTUNE (Sept. 26, 2014, 4:29 PM), http://fortune.com/2014/09/26/ello-social-network/.

[9] FEDERAL TRADE COMMISSION, DATA BROKERS A CALL FOR TRANSPARENCY AND ACCOUNTABILITY 4 (2014) [hereinafter FTC REPORT], *available at* http://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

[10] *Id.* at app. C-7; Grant Gross, *FTC to Consider Stricter Online Privacy Rules*, PCWORLD (Dec. 7, 2009, 12:20 PM), http://www.pcworld.com/article/183910/article.html (quoting FTC Chairman Jon Leibowitz: "How many consumers . . . have ever heard the names of the many ad networks that end up with their information in the process of targeting ads . . . [h]ow many people understand the networks' role?") (alterations in original).

[11] *See* Alexis C. Madrigal, *How Much is Your Data Worth? Mmm, Somewhere Between Half a Cent and $1,200*, THE ATLANTIC (Mar. 19, 2012, 3:18 PM), http://www.theatlantic.com/technology/archive/2012/03/how-much-is-your-data-worth-mmm-somewhere-between-half-a-cent-and-1-200/254730/ ("If, post-facto, you inform people that you're already using their data, they won't pay to stop you. But if you ask beforehand, they might.")

privacy.[12] Although these guidelines are not binding, reputable companies have a strong incentive to adhere to them because they know that consumer trust is gained over years and lost in only instants.[13] Less reputable companies and outright scammers gather data simply because they know it can be sold or used in phishing or identity-theft-schemes.[14] Thus far, the Federal Trade Commission ("FTC") has focused on enforcement of companies' self-imposed privacy policies and has yet to provide minimum requirements for those policies.[15] Continued concerns have led academics to question whether the time for self-regulation has passed.[16] This Comment considers how regulation similar to the Unfair, Deceptive, or Abusive Acts or Practices ("UDAAP") standard, which currently protects the customers of financial institutions, could be adapted to protect consumers online while allowing companies to leverage the valuable benefits that data can provide.

Part II gives a brief review of the benefits and risks created by the analytical power that big data can provide, current methods of online data collection and use, the legal framework currently in place, and why banking regulations would serve as a good model for the data industry. Part III gives an outline of risk-based, disclosure-oriented regulatory guidelines modeled after standards developed in the banking industry.

II. BACKGROUND

A. The Power of Big Data

The basic object of statistical analysis is to use samples to generalize about a larger population.[17] More data lends greater predictive power to the generalizations that result from the study.[18] Data gathered through the Internet not only provides access to a nearly endless supply of

---

[12] *See, e.g.*, DIRECT MARKETING ASSOCIATION, DIRECT MARKETING ASSOCIATION'S GUIDELINES FOR ETHICAL BUSINESS PRACTICE 20 (2014), *available at* https://thedma.org/wp-content/uploads/DMA-Ethics-Guidelines.pdf.

[13] Telephone Interview with Chuck Sharp, CEO, RightIntel (Sept. 15, 2014) [hereinafter Sharp Interview]; *see also* Kashmir Hill, *How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did*, FORBES (Feb. 16, 2012, 11:02 AM), http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/.

[14] Sharp Interview, *supra* note 13.

[15] Jonathan P. Cody, Comment, *Protecting Privacy Over the Internet: Has the Time Come to Abandon Self-Regulation?*, 48 CATH. U. L. REV. 1183, 1211 (1999), http://scholarship.law.edu/lawreview/vol48/iss4/7.

[16] Dustin D. Berger, *Balancing Consumer Privacy with Behavioral Targeting*, 27 SANTA CLARA COMPUTER & HIGH TECH. L.J. 3–5 (2010); Cynthia Dwork & Deirdre K. Mulligan, *It's Not Privacy and It's Not Fair*, 66 STAN L. REV. ONLINE 35, 36 (2013); Robert Sprague and Corey Ciocchetti, *Preserving Identities: Protecting Personal Identifying Information Through Enhanced Privacy Policies and Laws*, 19 ALBANY L. J. SCI. & TECH. 91, 91–92 (2009); Cody, *supra* note 15, at 1191–92; Ryan Moshell, . . . *And Then There Was One: The Outlook for a Self-Regulatory United States Amidst a Global Trend Toward Comprehensive Data Protection*, 37 TEX. TECH. L. REV. 357, 359 (2005).

[17] MICHAEL SULLIVAN, III, STATISTICS: INFORMED DECISIONS USING DATA 5 (Sally Yagan et al. eds., Student ed. 2004).

[18] *Id.* at 464–65.

data points, but also allows data to be gathered on an individual basis, eliminating the need for generalizations. Both the generalizations obtained through analysis of large datasets and the individual data itself carry risks and benefits for consumers.

1. Benefits and Risks of Aggregate Data

Increasing predictive power through increasing sample size is the driving force behind machine learning. Rather than programming the proper response to every problem an application might encounter, machine learning allows a computer program to gather data until it learns how to respond.[19] For example, in the 1950s an IBM engineer wrote a program that allowed him to play checkers against a computer.[20] Because the computer only understood the concept of a legal move, he could easily beat it.[21] After designing a process that allowed the computer to record the odds of winning given a board configuration, he was still able to beat it.[22] But after allowing the computer to play against itself, it was able to produce enough data to develop accurate probabilities of winning based on each board configuration.[23] This proceeded until Arthur Samuel could no longer win a single game against his computer opponent—he had designed a machine that was better than he was at the task he had taught it.[24]

Kenneth Cukier, the data editor for The Economist, noted that the same machine-learning processes that allowed a computer to learn checkers strategy could allow self-driving cars to learn the rules of the road or for vehicle warning systems to alert drivers to potentially dangerous conditions on the road without requiring us to explicitly handle programmatically each and every situation the program might encounter.[25] But rather than just one car providing data, we might imagine if the data were shared between cars.[26] The time it would take for a computer to learn the processes involved in driving would be far less, just as the process for learning checkers could have gone faster with multiple computers playing multiple games and sharing the results.

In the medical field, machine learning was applied to create an algorithm that would identify cancerous cells within a biopsy.[27] Through

---

[19] Kenneth Cukier, Remarks at TEDSalon, Berlin 2014: Why Big Data is Better Data (June 2014) (transcript available at https://www.ted.com/talks/kenneth_cukier_big_data_is_better_data/ transcript?language=en).

[20] *Id.*

[21] *Id.*

[22] *Id.*

[23] *Id.*

[24] *Id.*

[25] *Id.*

[26] *See* Damon Lavrinc, *Feds Will Require All New Vehicles to Talk to Each Other*, WIRED (Feb. 3, 2014, 1:59 PM), http://www.wired.com/2014/02/feds-v2v/.

[27] Cukier, *supra* note 19.

this process, the computer was able to learn twelve factors that it used to predict cancer, three of which were previously unknown.[28] The sheer amount of data allowed the algorithm to recognize traits that traditional medical research had not been able to identify.[29] The application of a similar processes using data from social networking sites linked with medical records could aid in measuring outcomes and conducting observational studies.[30] The data could also aid in identifying and notifying people at risk for certain diseases.[31]

Data can also allow businesses to operate more efficiently. According to a well-known anecdote among data scientists, through an analysis of supermarket customers' purchases, statisticians identified a trend of diapers and beer being purchased frequently in the same transaction.[32] As a result, stores began to place beer and diapers in closer proximity, or in a place where customers would have to pass on their way, to the checkout counter increasing beer sales. While the story itself may just be a thought experiment, Amazon put the principle in action through its "frequently purchased together" feature with wild success.[33] Target, on the other hand, was the subject of outrage when they used similar analysis to predict that a teenage girl was pregnant and sent her coupons for baby items before she had told her parents.[34]

These types of pattern recognition processes yield other insights into individual behavior and characteristics. A recent study in the Proceedings of the National Academies of Sciences studied users' Facebook "likes" to identify which "likes" were the strongest predictors of high intelligence.[35] Among the top five was a page dedicated to curly fries.[36] While it is very unlikely the fries themselves have any causative effect on intelligence, the researchers found that because a smart person had created this page, that

---

[28] *Id.*

[29] *Id.*

[30] FRANCISCO GRAJALES ET AL., SOCIAL NETWORKING SITES AND THE CONTINUOUSLY LEARNING HEALTH SYSTEM: A SURVEY 1–2 (2014), *available at* http://www.iom.edu/~/media/Files/Perspectives-Files/2014/Discussion-Papers/VSRT-PatientDataSharing.pdf.

[31] Niven R. Narain, *From Big Data to Actionable Data: Has Our Biology Failed Us, or Have We Failed to Use It?*, WIRED (Aug. 22, 2014, 9:52 AM), http://www.wired.com/2014/08/big-data-actionable-data-biology-failed-us-failed-use/; *see also* Karen Curtin et al., *Familial Risk of Childhood Cancer and Tumors in the Li-Fraumeni Spectrum in the Utah Population Database: Implications for Genetic Evaluation in Pediatric Practice*, 133 INT. J. CANCER 2444, 2452 (2013), *available at* http://healthcare.utah.edu/huntsmancancerinstitute/research/labs/schiffman/images/21_Familialriskofchildhoodcancer.pdf.

[32] *See* Yen-Liang Chen et al., *A Data Mining Approach for Retail Knowledge Discovery with Consideration of the Effect of Shelf-Space Adjacency on Sales*, 42 DECISION SUPPORT SYSTEMS 1503, 1504 (2006), *available at* http://www.sciencedirect.com/science/article/pii/S0167923606000030#.

[33] JP Mangalindan, *Amazon's Recommendation Secret*, FORTUNE (July 30, 2012, 11:09 AM), http://fortune.com/2012/07/30/amazons-recommendation-secret/.

[34] Hill, *supra* note 13.

[35] Michal Kosinski et al., *Private Traits and Attributes are Predictable from Digital Records of Human Behavior*, 110 PROC. NAT'L ACAD. SCI. 5802, 5804 (Apr. 9, 2013), *available at* http://www.pnas.org/content/110/15/5802.full.pdf+html.

[36] *Id.*

person was more likely to have smart friends and so it was the way the page propagated through the social network—not the underlying content—that made it predictive of higher intelligence.[37] These types of processes not only lend insight into the characteristics of individuals, but can also lend themselves to development of a kind of topology of how ideas and behaviors move through social networks.[38]

Using data from social media, researchers were able to create mechanisms to predict political preference, gender, sexual orientation, religion, and age.[39] At times, the content of the data itself directly reveals the attribute without the need for statistical analysis, for example, affiliation with gay groups like "Being Gay" or "Gay Marriage."[40] The researchers note that the process they use raises "negative implications, because it can easily be applied to large numbers of people without obtaining their individual consent and without them noticing . . . [and] [o]ne can imagine situations in which such predictions, even if incorrect, could pose a threat to an individual's well-being, freedom, or even life."[41]

The crux of the danger that big data presents is that consumers are largely unaware of the consequences of the seemingly mundane things that they share everyday through their online activities.[42] Theoretically, analysis of the same, publically available data that researchers have used to predict an individual's age, race and gender could also be used to predict drug and alcohol use, propensity for criminal activity, any number of hiring criteria, or even credit risk.[43] This also presents the risk of discrimination for protected classes because membership in that class could be predicted using data from social media and hidden within an algorithm, allowing organizations to hide discriminatory practices.[44]

2. Benefits and Risks of Individualized Data

Predictive generalizations are not the only application of large datasets. Data can have an impact on an individual basis. In her TED talk,

---

[37] Jennifer Golbeck, Remarks at TEDx, MidAtlantic 2013: The Curly Fry Conundrum: Why Social Media "Likes" Say More Than You Might Think (Oct. 2013) (transcript available at https://www.ted.com/talks/jennifer_golbeck_the_curly_fry_conundrum_why_social_media_likes_say_more_than_you_might_think/transcript?language=en (suggesting that homophily, the tendency for people to befriend those that are like themselves, might explain the relationship between high intelligence and curly fries).

[38] Eric Gilbert & Garrie Karahalios, *Predicting Tie Strength with Social Media*, CHI Proceedings of the 27th International Conference on Human Factors in Computing Systems 212 (2009), *available at* http://leonidzhukov.net/hse/2011/seminar/papers/chi09-tie-gilbert.pdf; Golbeck, *supra* note 37.

[39] Kosinski, *supra* note 35, at 5804.

[40] *Id.*

[41] *Id.* at 5805.

[42] *Id.* at 5802; Golbeck, *supra* note 37.

[43] Cukier, *supra* note 19.

[44] *See* Edith Ramirez, Chairwoman, Remarks on Behalf of the Federal Trade Commission: Big Data: A Tool for Inclusion or Exclusion? (Sept. 15, 2014) (transcript available at http://www.ftc.gov/news-events/audio-video/video/big-data-tool-inclusion-or-exclusion-part-1).

Talithia Williams shares the story of how tracking her individual data allowed her to make better decisions regarding her pregnancy.[45] Based on decades of data, doctors estimate due dates for pregnant women assuming a twenty-eight day menstruation cycle.[46] Because she kept track of her cycle, Ms. Williams knew that she did not fit this generalization, and was able to determine that the decision to induce labor was premature.[47] But, individualized data like this may already be available without requiring us to gather it ourselves.[48]

Both individualized and aggregate data allows marketers to target their advertisements, reducing the cost to the company as well as the annoyance to the potential consumer.[49] Marketers seek to identify a need and show the consumer how to fill it.[50] They don't want to market to uninterested consumers because it is expensive both in terms the cost of the advertisement as well as in terms of the potential business relationship.[51] Data allows marketing to work more efficiently for all parties.[52] Knowledge of a potential customer's individual likes and dislikes eliminates the guesswork associated with aggregate data.[53] A McKinsey report estimates the potential annual consumer surplus from using personal location data at $600 billion and the potential increase in retailers' operating margins at 60%.[54]

However, individualized data presents risks similar to those for aggregate data. Instead of the risk of being wrongly categorized, individual data has the potential to invade user privacy when it is collected without the user's knowledge, again, perhaps placing in jeopardy "an individual's well-being, freedom, or even life."[55] Moreover, the kinds of data now routinely collected could be used to perpetrate identity theft. One researcher found that using data available for purchase online allowed him to link a user's name to a full nine-digit social security number in 86% of his trials.[56] The 2012 Global Internet User Survey found that the unexpected use of personal

---

[45] Talithia Williams, Remarks at TEDxClaremontColleges: Own Your Body's Data (Feb. 2014) (transcript available at https://www.ted.com/talks/talithia_williams_own_your_body_s_data/transcript?language=en).

[46] *Id.*

[47] *Id.*

[48] *See* Kosinski, *supra* note 35, at 5802.

[49] *See* Saul Hansell, *A Guide to Google's New Privacy Controls*, N.Y. TIMES (Mar. 12, 2009, 6:08 AM), http://bits.blogs.nytimes.com/2009/03/12/a-guide-to-googles-new-privacy-controls/?_r=1.

[50] Personal Interview with Mike Brian, Partner, Penna Powers (Sept. 16, 2014).

[51] *Id.*

[52] *Id.*

[53] *See* Kosinski, *supra* note 35, at 5804–05.

[54] JAMES ET AL., MCKINSEY GLOBAL INST., BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION, AND PRODUCTIVITY 2 (2011), *available at* http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation; s*ee also* Rathburn & Dunlap, *supra* note 7.

[55] Kosinski, *supra* note 35, at 5805.

[56] David Goldman, *Rapleaf is Selling Your Identity*, CNN MONEY (Oct. 21, 2010, 1:13 PM), http://money.cnn.com/2010/10/21/technology/rapleaf/ (noting further that the researcher asked not to be identified because of business dealings with several companies in the field).

information disclosed online resulted in loss of money in 8% of cases, theft of personal information in 17% of cases, impersonation in 9% of cases, and credit damage in 4% of cases.[57]

B. Methods of Collection and Use

In addition to mining public records and the data that users knowingly share online, companies use technologies embedded in their websites to generate data on users. Aggregators purchase this data and use it for targeted marketing or serving personalized content in apps.

1. Methods of Online Data Collection

The familiar means of tracking user behavior online is through the use of cookies.[58] Upon logging into a service, a file containing a unique string of characters is installed on the user's computer.[59] This allows the site to recognize the user as they navigate the website without the need for reauthentication.[60] What consumers might not know is that this file also gives the site the ability to log the way they interact with the site.[61] Consumers will soon produce even more data as items like appliances and cars are added to the "Internet of Things."

Normally cookies only track interactions at a single website.[62] "Supercookies," as they have come to be called, have the ability to track a user across multiple websites and recreate cookies after a user deletes them.[63] After the practice was publicized in the media, user outcry caused most companies to disclaim use of supercookies.[64] Nevertheless, a new variety of supercookie has resurfaced among cell phone service providers using a similar method.[65]

Canvas fingerprinting is a new technology with the potential to replace supercookies.[66] The "canvas" tag in a new version of the computer

---

[57] *Global Internet User Survey 2012*, INTERNET SOCIETY, http://www.internetsociety.org/apps/surveyexplorer/online-privacy-and-identity/what-were-the-consequences-of-the-unexpected-use-of-personal-information-disclosed-online-14/ (last visited Feb. 15, 2015).

[58] *See* A.E.S., *How Online Advertisers Read Your Mind*, THE ECONOMIST (Sept. 21, 2014, 11:50 PM), http://www.economist.com/blogs/economist-explains/2014/09/economist-explains-12.

[59] *Id.*

[60] *Cookies: Frequently Asked Questions*, ABOUTCOOKIES.ORG, http://www.aboutcookies.org/default.aspx?page=5 (last visited January 26, 2015) [hereinafter *Cookies FAQ*].

[61] A.E.S., *supra* note 58.

[62] *See Cookies FAQ*, *supra* note 60 (describing session, or transient cookies and permanent cookies).

[63] *Id.* (describing flash cookies); *see also* Nicholas Jackson, *The Next Online Privacy Battle: Powerful Supercookies*, THE ATLANTIC (Aug. 18, 2011, 10:31 AM), http://www.theatlantic.com/technology/archive/2011/08/the-next-online-privacy-battle-powerful-supercookies/243800/.

[64] Jackson, *supra* note 63.

[65] Robert McMillan, *Verizon's "Perma-Cookie" is a Privacy-Killing Machine*, WIRED (Oct. 27, 2014, 6:30 AM), http://www.wired.com/2014/10/verizons-perma-cookie/.

[66] GUNES ACAR ET AL., THE WEB NEVER FORGETS: PERSISTENT TRACKING MECHANISMS IN THE WILD 676 (2014), *available at* https://securehomes.esat.kuleuven.be/~gacar/persistent/the_web_never_forgets.pdf.

language used to program web sites includes a feature that allows forms to be drawn on the fly using a script that translates unique characteristics of the user's computer into an image similar to a fingerprint.[67] Through a process of recognizing variations in the way the form is drawn and hiding it from the user, websites can recognize the user even after the user has deleted any cookies installed in their browser—all without notifying the user.[68] Even after using opt-out pages, researchers determined that the fingerprinting method continued collecting data.[69]

The Internet of Things is the result of adding Internet connectivity to more and more devices.[70] It will allow users "to digitally identify, observe and control objects in the physical world" via the Internet.[71] Everything from sprinkler systems[72] to kitchen appliances[73] is being converted into an Internet device. Data produced by these devices will be invaluable to companies seeking to gain insight into the way users interact with their products.[74]

2. Data Aggregation

Beyond using data for different kinds of analysis or marketing, a growing area of user concern is the sale of data to companies that combine data from several sources.[75] But not all of these "aggregators" appear to pose the same level of concern. To illustrate the variety of ways aggregators function, consider two examples: Rapleaf and Factual.

Rapleaf was one of several companies in the business of constructing consumer profiles using data gathered from social media.[76] They gathered email and physical addresses, income, and likes and dislikes, including music habits from Pandora and wish lists from Amazon.[77] The company had detailed profiles for over 400 million web users.[78] This data could be linked back to the individual user's Facebook account, providing several means of contacting the user.[79] The company offered a way for consumers to log in and view these profiles, but this offers little relief

---

[67] *Id.* at 674.
[68] *Id.* at 674–75.
[69] *Id.* at 685.
[70] JAN HÖLLER ET AL., FROM MACHINE-TO-MACHINE TO THE INTERNET OF THINGS: INTRODUCTION TO A NEW AGE OF INTELLIGENCE 3–4 (2014).
[71] *Id.* at 11.
[72] *E.g.* SKYDROP, http://www.skydrop.com/ (last visited Feb. 15, 2015).
[73] Michael Wolf, *How the Internet of Things is Reinventing the Kitchen,* FORBES (July 31, 2014, 3:35 PM), http://www.forbes.com/sites/michaelwolf/2014/07/31/how-the-internet-of-things-is-reinventing -the-kitchen/.
[74] JAN HÖLLER ET AL., *supra* note 70, at 65, 66.
[75] *See* FTC REPORT, *supra* note 9, at 4.
[76] *Id.* at 8–9.
[77] Goldman, *supra* note 56.
[78] *Id.*
[79] *Id.*

because few consumers even know they exist and the level of detail they provide.[80] As David Goldman from CNN Money noted, "[i]t's one thing for a marketer to know you're 40 years old and subscribe to travel magazines; it's another for them to know you're leaving Saturday for a week in Italy."[81] Moreover, because the data can be linked with other data sets, the information can be traced back to a particular individual, allowing the aggregator to create detailed profiles of individual users.[82] Malicious use of this data can result in serious consequences including loss of money, theft of personal information, impersonation, and credit damage.[83]

Factual created a hub of data to allow the user experience for smartphone apps to be personalized for each user based on the user's location.[84] Businesses provide Factual with data about the location of the business and services.[85] As a consumer uses an app at a place that Factual recognizes as a golf course, Factual might infer that the consumer is a golfer.[86] Gil Elbaz, Factual's founder, assures us that "[a]ll of this can be done without knowing anything about who they are, their email address, or anything that's privately identifiable . . . [w]e're simply in the background helping these companies make sense of the information they already have."[87] By aggregating information on businesses, Factual allows apps to provide personalized content without the need to gather information on users.

C. Current Legal Framework

The major regulator in the data industry is the FTC, but other regulators like the Consumer Financial Protection Bureau ("CFPB") also exercise a role in enforcing the patchwork of laws intended to protect data under different circumstances.

Federal laws regulate privacy using a "sectoral rather than omnibus approach."[88] The Gramm-Leach-Bliley Act ("GLBA") regulates financial data, requiring financial institutions to provide notice of their privacy policies and the opportunity to opt out of having information shared.[89] The Fair Credit Reporting Act ("FCRA") places limits on users and producers of

---

[80] *Id.*
[81] *Id.*
[82] *Id.*
[83] *Global Internet User Survey 2012*, *supra* note 57.
[84] Issie Lapowsky, *The Next Big Thing You Missed: How to Add Location Data to Your App Without Relying on Google*, WIRED (Sept. 9, 2014, 6:30 AM), http://www.wired.com/2014/09/factual/.
[85] *Id.*
[86] *Id.*
[87] *Id.*
[88] Woodrow Hartzog & Daniel J. Solove, *The Scope and Potential of FTC Data Protection*, 83 GEO. WASH. L. REV. (forthcoming 2015) (manuscript at 26), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2461096.
[89] 15 U.S.C. §§ 6801–6827 (2012).

consumer credit reports including imposing duties with regard to accuracy and dispute resolution.[90] The Health Insurance Portability and Accountability Act ("HIPPA") regulates medical data, setting standards for its use and protection.[91] The Controlling the Assault of Non-Solicited Pornography and Marketing Act ("CAN-SPAM") regulates collection and use of email addresses.[92] The Electronic Communications Privacy Act ("ECPA") regulates data from the content of electronic communications.[93]

The FTC is calling for greater transparency and accountability for data brokers, but has largely declined to provide hard and fast rules, leaving privacy to be regulated according the policies of each individual company.[94] While HIPPA and FCRA offer protections for health and financial data, there are no minimum standards for generic types of data other than those companies that choose to self-impose through their privacy policies.[95]

Some states have enacted data privacy laws to fill this void. California requires companies to disclose to consumers when they share personal information with third parties,[96] put reasonable security measures in place,[97] and provide users under eighteen-years old with a right similar to the European Union's ("EU") "right to be forgotten."[98]

D. Looking to the Banking Industry

In a Forrester survey, consumers ranked banks as the type of company they consider most trustworthy when it comes to their data with 45% of respondents indicating that they trusted banks.[99] With banking's reputation for honesty only beginning to recover after plummeting during the financial crisis in 2009 and 2010,[100] we might well ask how banks are

---

[90] *Id.* § 1681i(a)(8).

[91] 42 U.S.C. § 1320d et seq. (2012).

[92] 15 U.S.C. § 7701 (2012).

[93] 18 U.S.C. 2510 (2012).

[94] Hartzog & Solove, *supra* note 88, at 12; Steve Lohr, *New Curbs Sought on the Personal Data Industry*, N.Y. TIMES, May 28, 2014, at B1, *available at* http://www.nytimes.com/2014/05/28/technology/ftc-urges-legislation-to-shed-more-light-on-data-collection.html.

[95] Hartzog & Solove, *supra* note 88, at 19.

[96] CAL. CIV. CODE §§ 1798.83–.84 (West 2009 & Supp. 2015).

[97] *Id.* § 1798.81.5.

[98] CAL. BUS. & PROFESSIONS CODE §§ 22580–81. In light of data for services business model, this raises significant concerns for online companies. At its extreme, this law could allow consumers to receive a service and then demand to change the terms of the agreement under which those services were provided. Critics of this right in the EU equate it with "marching into a library and forcing it to pulp books." Rory Cellan-Jones, *EU Court Backs "Right to be Forgotten" in Google Case*, BBC (May 13, 2014, 8:40 PM), http://www.bbc.com/news/world-europe-27388289; *see also* Shaun A. Sparks, Comment, *The Direct Marketing Model and Virtual Identity: Why the United States Should Not Create Legislative Controls on the Use of Online Consumer Personal Data*, 18 Dick. J. Int'l L. 517, 540–41 (2000) (arguing that an opt-in mechanism for consumer data protection is an unconstitutional restriction of commercial speech).

[99] Eric Bloom, *New Forrester Study Shows Privacy Concerns and User Behavior are Inconsistent*, THE ONLINE PRIVACY BLOG (June 10, 2014), http://www.abine.com/blog/2014/forrester-privacy-study/.

[100] *See, e.g.*, *Honesty/Ethics in Professions,* GALLUP, http://www.gallup.com/poll/1654/honesty-ethics-professions.aspx#1 (last visited Feb. 15, 2015).

gaining trust.

First, banks are required to make initial disclosures before allowing electronic transfers to occur[101] and must provide an annual notice describing changes in terms and methods of resolving errors.[102] The content is standardized and even the formatting must adhere to a set of clear guidelines provided by regulators.[103] Second, the Dodd-Frank Wall Street Reform and Consumer Protection Act ("Dodd-Frank") requires financial institutions to refrain from any unfair, deceptive, or abusive acts or practices.[104] These UDAAP standards have allowed regulators to hold financial institutions accountable.[105] Unfair, deceptive, and abusive acts or practices are each associated with a separate legal standard.[106] The Consumer Financial Protection Bureau ("CFPB") noted that their understanding of the "unfair" and "deceptive" was informed by the FTC's understanding of those terms in the FTC Act.[107] Ironically, the FTC could now look to the CFPB to expand their enforcement regime beyond companies' self-imposed privacy policies. Indeed, data collectors and users could benefit substantially from a set of similar standardized disclosures and enforceable guidelines for the industry.

III. ANALYSIS

The ubiquity of data and its use across industries and geographic boundaries call for a standardized approach to protecting privacy. The system currently in place leaves consumers as well as companies relying on unenforceable industry best practices and guidelines. The FTC enforces the privacy policies that companies choose to provide, but each business has a different privacy policy so consumers are forced to either understand the complicated provisions on each website or risk forfeiting their privacy. A standardized set of minimum requirements would bring certainty to the market for both businesses and consumers.[108]

In order to ensure that the exchange of access to online services in

---

[101] 12 C.F.R. § 1005.7 (2014).

[102] *Id.* § 1005.8(b).

[103] *See generally id.* at app. A to pt. 1005.

[104] *See generally* 12 U.S.C. §§ 5481, 5531, 5536(a)(1)(B) (2012).

[105] *See* Chris Cumming, *Urban Trust Stops Charging Overdraft Fees on Prepaid Cards*, AMERICAN BANKER (Feb. 4, 2014), http://www.americanbanker.com/issues/178_24/urban-trust-stops-charging-overdraft-fees-on-prepaid-cards-1056430-1.html; Steven Forry, *2012: The CFPB Set Its Sights on Credit Card Companies*, A.B.A. (Mar. 22, 2013), http://apps.americanbar.org/buslaw/blt/content/2013/03/article-02-forry.shtml.

[106] UDAAP Bulletin, Consumer Fin. Protection Bureau, Prohibition of Unfair, Deceptive, or Abusive Acts or Practices in the Collection of Consumer Debts (July 10, 2013) [hereinafter UDAAP Bulletin], *available at* http://files.consumerfinance.gov/f/201307_cfpb_bulletin_unfair-deceptive-abusive-practices.pdf.

[107] *Id.* at 1 n.1.

[108] *See* Hartzog & Solove, *supra* note 88, at 40–41, 46; Rathburn & Dunlap *supra* note 7. The FTC would be able to initiate regulatory action without legislation, but additional rulemaking authority and greater jurisdiction over non-commercial entities that engage in commercial practices might make the regulatory scheme more effective. *See id.*

return for consumer data is meaningful and fair, the consumer must, at the very least, be kept informed of the actual price they are paying in exchange for the services they use online. But, regulations need not go so far as those that govern health and financial data. Less sensitive data warrants less regulatory scrutiny. However, there should be a floor to the privacy afforded to users and rules to bring uniformity to disclosures. Regulations currently in place for financial institutions could serve as a model to produce a risk-based, disclosure-oriented regulatory scheme for the data industry.

A. Adapting the Standards Based on User-Site Relationship, Information Sensitivity, and Use

As a preliminary matter, the rules governing data should adapt based on the wide variety of relationships between sites and users, data types, and data uses. The Organization for Economic Cooperation and Development's benchmark for fair information collection begins by recommending disclosure of the purpose of the data collection and obtaining consent before data is collected.[109] In the 2012 version, it noted that prior affirmative consent in all cases would be impractical.[110] This makes especially good sense in cases where the relative risk to consumers is diminished. The FTC could consider the following as factors that play into determining the level of risk to the consumer: (1) whether the user sets up an account, (2) the type of data collected, and (3) whether data is sold to or shared with third parties.

When a consumer sets up an account, it suggests the expectation of a permanent relationship. A user should, therefore, expect the website to gather more information as a part of the terms of use after appropriate disclosures are provided.[111] But, when the consumer chooses not to set up an account, the assumption is that the relationship is only temporary. The account signup process gives users the opportunity to consider whether, and to what extent, to allow the site to track their activities—provided they are presented with appropriate disclosures as a part of that process. Under these conditions, the risk that consumers are being tracked unwittingly is greatly diminished and the consumer can then be expected to manage what they choose to reveal. Requiring an account signup when the site conducts persistent tracking methods thus diminishes the level of risk to the

---

[109] ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, OECD GUIDELINES ON THE PROTECTION OF PRIVACY AND TRANSBORDER FLOWS OF PERSONAL DATA 14 (2013), *available at* http://www.oecd.org/sti/ieconomy/2013-oecd-privacy-guidelines.pdf.

[110] ORGANIZATION FOR ECONOMIC COOPERATION AND DEVELOPMENT, OECD GUIDELINES ON THE PROTECTION OF PRIVACY AND TRANSBORDER FLOWS OF PERSONAL DATA (2012), *available at* http://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm ("The knowledge or consent of the data subject is as a rule essential, knowledge being the minimum requirement. On the other hand . . . there are situations where for practical or policy reasons the data subject's knowledge or consent cannot be considered necessary.").

[111] *See infra* Part III.B.1.

consumer.

Likewise, collection of certain kinds of data pose more risk to consumer privacy than others. Knowing that a particular consumer is a golfer or reads articles about constitutional law presents a low level of risk. Data that includes methods to contact the individual or information that allows the data to be linked with other datasets presents a greater risk to consumers.[112] Analysis of non-identifying data yields generalizations about different types of consumers, while analysis of personally-identifying-data ("PID" or "PII") allows greater access into the personal details of an individual's life.

Users decide which sites they will visit and what information they will share. Selling data to other sites presents a greater risk of invading privacy because the user no longer has control over which sites have their data. The fact that a site sells or shares data presents a greater risk to consumer privacy because the consumer loses control of the data after it is sold.

The regulatory requirements should consider these factors when determining which requirements to impose and the extent of the requirements. For example, the fact that a site sells or shares data would tend toward requiring prior, affirmative disclosure, but if the site were only sharing non-identifying information, the diminished risk might make a link to the policy online sufficient.

B. Potential Regulatory Requirements

Regulations should provide uniform rules on disclosures that alert consumers to the scope of the data collection and use, when consumers must be given the opportunity to opt-out of collection and stop the use, and limits on the kinds of data that can be collected based on the services provided.

1. Disclosures

Any regulatory system for data gathering should focus on disclosure. Surveys show that consumers are not upset by the fact that they are being tracked, indeed many enjoy the benefits.[113] It is the fact that data is being gathered without their knowledge that poses the problem.[114]

Upon signing up to use a website, users, perhaps unwittingly, consent to an exchange of the site's services for their personal information.[115] In order to make an informed decision, consumers must be

---

[112] *See supra* Part II.B.2.

[113] Bloom, *supra* note 99.

[114] *See* Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN L. REV. ONLINE 41, 41–43 (2013).

[115] Gross, *supra* note 10.

alerted to the terms of this exchange. Advocates for the self-regulatory scheme currently in place argue that privacy policies are sufficient to alert users to this payment.[116] Since there are no minimum requirements for these policies, users must bear the transactional costs in time and effort to understand the policies of each site they use or risk forfeiting their privacy.

Before allowing consumers to set up electronic transfers or apply for card accounts, financial institutions must provide consumers with disclosures substantially in the same form as the model forms provided in the rule.[117] Recently, the CFPB began testing a new form for prepaid card disclosures after finding that the varying disclosures produced by each company made it difficult for the consumer to make comparisons.[118] Likewise, a standardized disclosure provided as a part of the signup process for a website could disclose its privacy practices regarding types of information it tracks and the way the information can be used. For example, consider Capital One's privacy statement,[119] which includes a list of the personal information collected and a box that lists how it uses personal information and discloses the opportunity to opt out. This includes whether they sell data to third parties.

As argued above, the regulatory scheme should be adjusted based on the level of privacy risk. Due to the sensitive nature of financial data, the GLBA requires disclosures on a yearly basis.[120] For more mundane types of data, disclosure upon signup with access to revoke consent through a settings page is probably sufficient. If a site does not gather personally identifying information, even less would be required—a notification on the page with a link to the site's privacy policy is probably sufficient.[121]

2. Opt-Out and Revocation of Consent

In banking regulations, one concern that has arisen with requiring banks to obtain consent from customers before providing overdraft services is whether banks will attempt to coerce consumers into opting into the service.[122] Regulations require banks to offer the same terms, conditions, and features to consumers that opt out.[123] This is perhaps because regulators

---

[116] *See, e.g.*, Sparks, *supra* note 98, at 539.

[117] 12 C.F.R. § 1005.7 (2014).

[118] Eric Goldberg, *Prepaid Cards: Help Design a New Disclosure*, CONSUMER FIN. PROT. BUREAU (Mar. 18, 2014), http://www.consumerfinance.gov/blog/prepaid-cards-help-design-a-new-disclosure/.

[119] *What Does Capital One Do with Your Personal Information?*, CAPITALONE, http://www.capitalone.com/media/doc/corporate/english-privacy-notice.pdf (last visited Feb. 15, 2015) [hereinafter *Capital One Disclosure*].

[120] 15 U.S.C. § 6803 (a) (2012).

[121] *See, e.g.*, *Cookies Info*, THE ECONOMIST, http://www.economist.com/cookies-info (last visited Feb. 15, 2015). At the time of writing, the Economist provided a link to this page in a header that also states that, by using the site, the user consents to the cookie policy.

[122] *See* FED. DEPOSIT INS. CORP., FDIC OVERDRAFT PAYMENT SUPERVISORY GUIDANCE (Nov. 24, 2010), *available at* https://www.fdic.gov/news/news/financial/2010/fil10081b.pdf

[123] 12 C.F.R § 1005.17 (2014).

take the position that overdraft programs are inherently harmful to consumers.[124]

Capital One's privacy statement explains that the consumer cannot opt out of every type of data use.[125] Likewise, websites provide services based on the assumption that they will be able to benefit from the user's data.[126] If a user were given power to retroactively withdraw consent, some sites would not be able to provide the service without charging a fee.[127] A site should be able to decide whether to offer its services on a fee basis or simply deny services. Companies should not, however, have the ability to attract users using a favorable privacy policy and then coerce them to opting in to data sharing to continue receiving the service. Likewise, users should not be able to use a "right to be forgotten" to change the terms of service after the fact.[128]

3. Usage Based Data Collection

Some websites gather much more information about users than is necessary to allow the site to function.[129] For example, the FTC issued a consent order to the makers of a flashlight app that gathered and transmitted "their devices' precise geolocation along with persistent device identifiers to various third parties."[130] The FTC focused on the fact that end user license agreement ("EULA") misrepresented that consumers had the right to opt out of this data collection, labeling this an unfair and deceptive practice.[131] The problem with this practice goes much deeper than any misrepresentation in the EULA.

The HIPPA requires entities to limit the use and disclosure of protected health information to the minimum necessary to accomplish the intended purpose.[132] Likewise, when a consumer agrees to allow a company to collect their data in exchange for the use of that company's services, there is an expectation that the data gathered will be related to—or at the very least commensurate with—that company's services. For example, a user would expect a mapping program to track location data, but the user would not expect a flashlight app to track their location. In order to keep the exchange between users and websites fair, data collection must be

---

[124] *See* G. Michael Flores & Todd Zywicki, *Overdraft Protection and Consumer Protection: A Critique of the CFPB's Analysis of Overdraft Programs*, 33 No. 3 Banking & Fin. Services Pol'y Rep. 10, 13 (2013).

[125] *Capital One Disclosure*, *supra* note 119.

[126] Zuckerman, *supra* note 3.

[127] *See id.*

[128] *See* Cellan-Jones, *supra* note 98.

[129] *See* Goldman, *supra* note 56.

[130] Goldenshores Technologies, LLC and Erik M. Geidl; Analysis of Proposed Consent Order to Aid Public Comment, 78 Fed. Reg. 75350, 75351 (Dec. 11, 2013) [hereinafter Goldenshores].

[131] *Id.*

[132] 45 C.F.R. §§ 164.502(b), 164.514(d)(3)(iii)(A) (2013).

commensurate with the services rendered to the user.

C. Enforcement

1. UDAAP as the Measuring Stick

UDAAP standards developed in the banking industry could aid the FTC in enforcing the standards it promulgates. Until now, FTC actions have focused on enforcing businesses' self-imposed privacy policies, arguing unfairness and deception when the businesses fail to do so.[133] UDAAP could offer a more robust set of guidelines for data collection practices.

Under UDAAP, a practice is unfair when (1) it causes injury, (2) the injury is not reasonably avoidable, and (3) the injury is not outweighed by a countervailing benefit to the consumer or to competition.[134] A practice is deceptive when it (1) misleads or is likely to mislead a consumer, (2) the consumer's misinterpretation would be reasonable under the circumstances, and (3) the deception is material to the "consumer's choice of, or conduct regarding, the product or service."[135] Abusive practices interfere with the consumer's ability to understand a term or condition of the service or takes unreasonable advantage of a consumer's (1) lack of understanding regarding risks, costs, or conditions; (2) inability to protect that consumer's best interests; or (3) reasonable reliance on the business to act in the consumer's best interests.[136] With few, minor changes, these standards could serve as the standards for the data collection and analysis industries.

The injury requirement in the unfairness standard usually focuses on the loss of money. Here, the currency involved in the exchange is the user's data.[137] Unfairness in data collection and use should therefore focus on the opportunity for the consumer to avoid being tracked and whether the data gathered is in proportion with the services provided. For example, failing to disclose the data that will be tracked as a result of using the site, or setting up an account, is unfair because the consumer is unwittingly giving away their data. Likewise, a service's data gathering is also unfair when it is not associated with, or proportional to, the benefit that the user receives. For example, a flashlight app's gathering of location data is unfair because it fails to provide a commensurate benefit in exchange for that data.[138]

Deception occurs when a user is misled in a material way with regard to the exchange taking place. The FTC already enforces this part of

---

[133] *See* Goldenshores, *supra* note 130, at 75351.
[134] UDAAP Bulletin, *supra* note 106, at 2.
[135] *Id.* at 3–4.
[136] *Id.* at 4.
[137] Zuckerman, *supra* note 3.
[138] *See* Goldenshores, *supra* note 130, at 75351.

UDAAP as it evaluates EULAs and takes action when companies fail to adhere to them.[139] But, this standard could also extend to the form of the disclosures that would be required. Consumers should be able to assume that sensitive data cannot be collected without prior, and effective, disclosure. Depending on the relative sensitivity of the data, if a disclosure fails to alert consumers that they are being tracked, the disclosure might be considered not only unfair, but potentially deceptive because the user is being materially misled as to the price that they are actually paying in exchange for the services.

The abuse standard under UDAAP has been the most troublesome for the financial industry,[140] but it need not pose the same problems in the data industry. Abuse occurs when the business takes unreasonable advantage of the user's relative position with regard to understanding the costs and risks associated with the exchange. One of the common definitions of abuse is "improper or excessive use . . . ."[141] In the data collection industry, abuse could therefore focus on improper uses of the data rather than abuse of the user providing it. A data company's failure to adequately secure data or use of the data to harm a consumer's credit (without first disclosing that the data gathered may be used in this way at the time of collection) could each be considered abusive.

2. Data Broker Registry

To aid enforcement, the FTC could also consider creating a registry for companies that purchase and aggregate data. Because of the risk of identity theft and the distance between consumer and data holder, a regulator is needed to ensure consumers are protected.

Federal law allows states to impose registration requirements on businesses that receive funds from consumers for the purposes of transferring the funds by wire, check, draft, facsimile, courier, etc.[142] The purpose of the registration is to aid investigation of money laundering and ensure that consumers are protected from financial loss. Similarly, the aggregation of personally identifying information lends itself to the risk of identity theft. Safety and soundness requirements should be imposed to mitigate the risk of identity theft because the consumer has no direct relationship with these businesses and therefore, has no control over their actions.

---

[139] *Id.*

[140] *See* Martin J. Bishop & Rebecca Hanson, *UDAAP: Shouldn't the CFPB Give Us More Information?*, 6 PAYBEFORE 1, 1–2 (2012), *available at* http://www.foley.com/files/Publication/660770e8-d671-4fcd-8018-4655f662dbcf/Presentation/PublicationAttachment/d1d29143-de1e-499c-9645-4721ea9c5c2c/RE_LE_v6_15.pdf.

[141] *Abuse Definition*, MERRIAM-WEBSTER, http://www.merriam-webster.com/dictionary/abuse (last visited Feb. 15, 2015).

[142] 18 U.S.C. § 1960 (2012).

IV. CONCLUSION

Even after the financial crisis, financial institutions are still the most trusted entities with regard to personal data. Regulations that have proven effective in the financial industry can be adapted into a risk-based, disclosure-oriented regulatory framework that will protect consumers while avoiding undue burden on Internet companies. The requirements under this framework should adapt based on the level of risk that a practice presents to consumers and should include disclosure and opt-out requirements as well as restrictions on data collection based on the service the data collector provides in exchange. UDAAP can serve as an effective measure for the FTC to apply as it enforces these requirements.