3-1-2017

Privacy by Design: Taking Ctrl of Big Data

Eric Everson
Herzing University

PRIVACY BY DESIGN: TAKING CTRL OF BIG DATA

ERIC EVERSON*

ABSTRACT

The concept of Privacy by Design is rooted in systems engineering. Yet, it is the legal framework of global privacy that gives new color to this concept as applied to Big Data. Increasingly, the long arm of the law is reaching into Big Data, but it is not simply by matter of regulatory enforcement or civil legal developments that Privacy by Design (PbD) is being thrust into the spotlight once more.

Given that Big Data is considered miniscule in contrast to future data environments,[1] PbD is simply the right thing to do. This paper aims to explore the origin of PbD, the current and future state of Big Data and regulatory enforcement, and the methodology of PbD applied to Big Data. As a cornerstone of organizational culture, PbD is a concept that allows organizations of any size to embrace the privacy interests of the data they collect, store, and use at the forefront of their approach.[2]

CONTENTS

* JD, MBA, MSIT-SE, Associate Faculty of Information Security, Herzing University. Mr. Everson is a technology attorney licensed by the Florida Bar. His work focuses on the intersection of technology, business, and the law. Areas of focus in his practice include privacy, bank regulation, financial technology, cyber and information security, social media, and intellectual property

[1] Ron Miller, *If You Think Big Data's Big Now, Just Wait*, TECH CRUNCH (Aug. 10, 2014), https://techcrunch.com/2014/08/10/big-data-bound-to-get-really-really-big-with-the-internet-of-things/.

[2] Peter Schaar, *Privacy by Design*, 3 IDENTITY INFO. SOC'Y 267, 267 (2010).

I. INTRODUCTION

Big Data notably has been referred to as the rocket fuel of economic growth.[3] As the field of big data progresses, maturity will develop as the focus moves away from the initial excitement that we can process large data and toward understanding the acquiring, stewarding, and sharing of our data.[4]

Turning to the world's foremost collection of aggregate data, Google's definition of "Big Data" is "[e]xtremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions."[5] So, with at least a baseline for why we value Big Data, the central theme of this paper is focused on leveraging the PbD framework for the purpose of taking control of this valuable asset of Big Data in its collection, storage, and use.

II. WHAT IS PRIVACY BY DESIGN?

To best understand the PbD framework, it should be noted that the concept is an evolving framework that was first applied to systems engineering.[6] Also, PbD has notable thematic applicability to the continual advancement of data collection, storage, and use.[7] PbD is a foundational approach that takes privacy into account at the forefront of the engineering lifecycle by culturally perpetuating privacy at all levels of an organization.[8] Continued refinement of PbD has yielded seven core tenants called the foundational principles, which include: 1) proactive not reactive, preventative not remedial; 2) privacy as the default setting; 3) privacy embedded into design; 4) full functionality—positive-sum, not zero-sum; 5) end-to-end security—full lifecycle protection; 6) visibility and transparency—keep it open; and 7) respect for user privacy—keep it user-centric.[9] These tenants will be explored in greater detail as this paper later examines the application of methodology to Big Data.

As a pedagogical framework, PbD encourages managers and creators to think about the data and privacy interests therein that are to be ingested at the forefront of the design process as opposed to being an afterthought in the development lifecycle.[10] PbD allows creators to specially architect environments and systems with considerations of data use for implementation at the onset, which will directly tie to business or operational processes once the solution is promoted into a live

---

[3] Edd Wilder-James, *Making a Moonshot? Put Data in Your Rocket*, FORBES (June 21, 2013), http://www.forbes.com/sites/edddumbill/2013/06/21/making-a-moonshot-put-data-in-your-rocket/.

[4] *Id.*

[5] *Big Data*, GOOGLE.COM, https://www.google.com/#q=definition+big+data (last visited Sept. 18, 2016).

[6] Peter Hustinx, *Privacy by Design: Delivering the Promises*, 3 IDENTITY INFO. SOC'Y 253, 253-54 (2010).

[7] Ann Cavoukian, *Privacy by Design: The 7 Foundational Principles*, IAB.ORG (2009), https://www.iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf.

[8] *Id.*

[9] *Id.*

[10] *Id.*

production status.[11] In view of rapid and dramatic technological change, it is important to take the special requirements of privacy protection into account early on because new technological systems often contain hidden dangers that are very difficult to overcome after the basic design has been worked out.[12]

The risk of neglecting the PbD approach is creating a solution or, worse yet, a data management culture rich in Highly Confidential data elements or Personally Identifiable Information (PII) attributes with a limited controls framework.[13] Historically, such approaches have notoriously given rise to future "bolt on" developments to address fundamental privacy vulnerabilities at the expense of time, money, reputation, security, or degradation of performance.[14] As a harbinger, the state of regulatory case law demonstrates, "A company does not act equitably when it publishes a privacy policy to attract customers who are concerned about data privacy, fails to make good on that promise by investing inadequate resources in cybersecurity, exposes its unsuspecting customers to substantial financial injury, and retains the profits of their business."[15]

Adopting a PbD approach allows organizations to consider the privacy interest of the data that a system or environment will collect, use, or store.[16] As a foundational principle, PbD is especially focused on protecting confidential personal and financial information, including full legal names, addresses, bank account data, social security numbers, and dates of birth.[17] Although serving as a contextual example, PbD is not limited to mere consideration of these data elements. Other notable considerations in the academic progeny of PbD recognize fair information practices common in most privacy legislation in use today: notice, choice and consent, proximity and locality, anonymity and pseudonymity, security, and access and recourse.[18]

At its core, PbD is an evolving framework with applicability to the continual advancement of data collection, storage, and use.[19] Although the PbD approach has only recently emerged, its prevalence in the continued development of data-rich environments is imminent as organizations continue to face new data use opportunities and challenges.[20] The design and implementation of privacy requirements in systems is a difficult problem and requires the translation of complex social, legal, and ethical concerns into systems requirements. The concept

[11] *See generally* Schaar, *supra* note 2.

[12] *See id.* at 274.

[13] *See generally* Paul M. Schwartz, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, N.Y.U. L. REV. 1814 (2011).

[14] *See* Marc Langheinrich, *Privacy by Design—Principles of Privacy-Aware Ubiquitous Systems*, 2201 UBICOMP 2001: UBIQUITOUS COMPUTING 273, 291 (2001); *see also* Schaar, *supra* note 2; Cavoukian, *supra* note 7.

[15] FTC v. Wyndham Worldwide Corp., 799 F.3d 236, 245 (3d Cir. 2015).

[16] Schaar, *supra* note 2.

[17] Storm v. Paytime, Inc., 90 F. Supp. 3d 359, 363 (M.D. Pa. 2015).

[18] Langheinrich, *supra* note 14, at 273.

[19] Jeroen van Rest et al., *Designing Privacy-by-Design*, *in* PRIVACY TECHNOLOGIES AND POLICY 55, 56 (Bart Preneel & Demosthenes Ikonomou eds., 2014).

[20] Hustinx, *supra* note 6, at 253.

of PbD has been proposed to serve as a guideline on how to address these concerns.[21] The opportunity that PbD introduces is the ability to foster a privacy-first culture that extends from organizational governance and leadership to design concepts that define brand trust.

III. PRIVACY BY DESIGN APPLIED TO BIG DATA

The Big Data environment creates unique challenges that the foundational principles of PbD help solve. What we are seeing as the constructs of Big Data evolve is that the traditional notions of protecting privacy via basic precepts such as de-identification and removal of PII data elements, while still foundational, are not enough in isolation.[22] To establish an effective Big Data approach, the foundational principles of PbD allow us to step back and consider the holistic lifecycle of the Big Data environments and data uses we employ.[23]

As we look to explore the opportunities of PbD in Big Data, once more, the foundational principles of PbD include: 1) proactive not reactive, preventative not remedial; 2) privacy as the default setting; 3) privacy embedded into design; 4) full functionality—positive-sum, not zero-sum; 5) end-to-end security—full lifecycle protection; 6) visibility and transparency—keep it open; and 7) respect for user privacy—keep it user-centric.[24] It is through the examination of each of these core tenants that the applicability of PbD to this emerging area is best considered.

A. Proactive Not Reactive; Preventative Not Remedial

Within a Big Data environment, or even at a campaign level, there is perhaps nothing more gut-wrenching than realizing that the data in a set, compilation, or derivative output has very real privacy implications. Harkening back to the classic tale of Target learning about a teenage pregnancy before the girl's father was informed,[25] we find data modeling methods violate sometimes-alarming, unintended privacy interests. In this instance, the Target statistician was not focused on the privacy interest of teen pregnancy, but rather, more general pregnancy related purchase trends that were conducted by assigning a generic de-identified "Guest ID" and analyzing pregnancy related purchase habits of consumers via transaction data.[26] This seemingly innocuous exercise in marketing data rapidly became a reputational

---

[21] Seda Gurses et al., *Engineering Privacy by Design*, PRIVACY AND DATA PROTECTION (2011), https://securewww.esat.kuleuven.be/cosic/publications/article-1542.pdf.

[22] *Data De-Identification: An Overview of Basic Terms*, PRIVACY TECH. ASSISTANCE CTR. (2013), http://ptac.ed.gov/sites/default/files/data/deidentification/terms.pdf [hereinafter *Data De-Identification*].

[23] Cavoukian, *supra* note 7.

[24] *Id.*

[25] Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES (Feb. 16, 2012), http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html (discussing Target's ability to identify pregnant shoppers through the collection of "vast amounts of data on every person who regularly walks into one of its stores").

[26] *Id.*

risk for the Target brand as outraged consumers and privacy professionals alike received the news.[27]

This classic tale reminds us that we should proactively consider the implications and perceptions that can result from the data we collect, store, and use. It is with consideration of the broader context that data might be examined that we should give pause to the preventative measures and assessment of data from the onset.

B. Privacy as the Default Setting

At present, the most popular Big Data tools do not include strong, automated safeguards related to post-contextual analysis of output data.[28] Although data environments can be set up to require de-identification of data at an element level on the front end,[29] as the Target example provides, it can be the way the data is modeled that yields alarming results through the lens of privacy.[30] Additionally, the sensitivity of data at all stages of the data lifecycle should always be considered at the onset.

Open source and non-native system of record data can be especially vulnerable to privacy risks because the full contents are potentially not under the same set of controls prior to introduction into the data environment.[31] Some of this risk can be mitigated via the Extraction-Transformation-Loading (ETL) tools, which are pieces of software responsible for the extraction of data from several sources, as well as their cleansing, customization, and insertion into a data warehouse.[32]

While tooling and approaches are continually being developed to mitigate privacy risks, the importance of developing a solid privacy culture within a data-driven organization is paramount. This includes strong public and internal privacy policies coupled with management's prioritization and enforcement of privacy governance.

C. Privacy Embedded Into Design

In today's operational environment, Big Data architects rarely are afforded the luxury of imagining their environments as a fresh build from the ground up. Unlike traditional transaction records collected from various legacy systems of the 1980s, the data that e-commerce systems collect from the web are less structured and often contain rich customer opinion and behavioral information.[33] These unique

---

[27] *Id.*

[28] Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, N.W. J. TECH. & INTELL. PROP. 240 (2013).

[29] Marit Hansen et al., *The Open Source Approach—Opportunities and Limitations with Respect to Security and Privacy*, COMPUTERS & SECURITY 461, 461-71 (2001).

[30] Duhigg, *supra* note 25.

[31] Josh Lerner & Jean Tirole, *The Economics of Technology Sharing: Open Source and Beyond*, J. ECON. PERSP. 99, 99-120 (2005).

[32] Panos Vassiliadis et al., *On the Logical Modeling of ETL Processes*, *in* ADVANCED INFO. SYS. ENG'G 782, 786 (A. Banks Pidduck et al. eds., 2002).

[33] Hsinchun Chen et al., *Business Intelligence and Analytics: From Big Data to Big Impact*, 36 MIS Q. 1165, 1169 (2012).

evolutions in the availability and character of the data introduce unique design opportunities, whether architecting from the ground up or from an existing platform.

By its most basic prima facie description, PbD literally describes privacy at the forefront of design when applied to a Big Data environment.[34] Whether facing the unique challenges of parsing transaction data from existing environments, intelligence extraction, opinion mining, question answering, topic-centric web mining, or social network analysis, there exist opportunities to frontload privacy considerations and controls into the design.[35] Here, PbD may be extended by establishing a privacy self-assessment that requires management and creators to more fully consider the context of their data through the lens of privacy and unique privacy interests inherent in their designs. Design here, being the convention for the construction of a data environment or campaign, requires the creator to think about the build plan and lifecycle with privacy in mind.

D. Full Functionality—Positive-Sum, Not Zero-Sum

When designing with privacy in mind, compromise is the term that perhaps comes up more often than not. This tenant of PbD reminds us to consider the win-win approach to building privacy conscious, data-rich environments or solutions.

With consideration of the more popularized anonymity and pseudonymity precepts that exist within privacy today, this allows us to contemplate such positive-sum functionality when related to something like a surrogate data element.[36] With regard to to the Target example,[37] this can be something as simple as a "GuestID" which acts as a de-identified element to reduce the privacy risk in a data set. The positive-sum often comes from the ability in such a case to build upon surrogate elements without introducing unnecessary privacy risk. So, in turn, by instituting a surrogate element, the privacy interest wins while creating a victory for the creator who has a new data element from which to build.

E. End-to-End Security—Full Lifecycle Protection

Perhaps we have all been told that if you do not want something to be publicly known, do not put it on the Internet. This advice may be especially true in the modern dawn of social media; yet, even in a public forum, there still exist certain expectations of privacy.

A recent example of privacy expectations in a public forum played out when a Facebook engineer accessed a Facebook user's profile after previously receiving the user's permission; the access occurred without the user providing his password.[38]

---

[34] Ira S. Rubinstein, *Regulating Privacy by Design*, 26 BERKELEY TECH. L.J. 1409, 1410 (2011).

[35] Chen et al., *supra* note 33, at 1170.

[36] George Tomko, *SmartData: The Need, the Goal and the Challenge*, *in* SMARTDATA: PRIVACY MEETS EVOLUTIONARY ROBOTICS 11, 12 (Inman Harvey et al. eds., 2013).

[37] Duhigg, *supra* note 25, at *1.

[38] Emil Protalinski, *Facebook Explains When Employees Can Access Your Account Without Your Password*, VENTUREBEAT (Feb. 27, 2015), http://venturebeat.com/2015/02/27/facebook-explains-when-employees-can-access-your-account-without-your-password/.

This prompted the media to inquire about when exactly the company's employees can perform such actions.[39]

Throughout the technology industry, Facebook generally is known to have strict user data access policies. Thus, it was not surprising that its response included statements such as,

> We have rigorous administrative, physical, and technical controls in place to restrict employee access to user data. . . . Access is tiered and limited by job function, and designated employees may only access the amount of information that's necessary to carry out their job responsibilities, such as responding to bug reports or account support inquiries. Two separate systems are in place to detect suspicious patterns of behavior, and these systems produce reports once per week which are reviewed by two independent security teams. We have a zero tolerance approach to abuse, and improper behavior results in termination.[40]

This response from Facebook demonstrates that, even regarding the user profiles which potentially (depending on the users' self-enabled privacy settings) could be totally open to public view, the company has deployed end-to-end automated safeguards to protect the privacy interest of their users.[41] Such best practices in user data privacy protection should not be isolated to data-driven businesses like Facebook but should emerge as common practice with full lifecycle protection considered in design regardless of industry or agency.

F. Visibility and Transparency – Keep it Open

Unique privacy risks can arise when obscuring visibility and transparency around source data. Although cloud-computing has emerged with great benefit, many companies and agencies alike have identified their concerns with transparency and visibility in such data environments as a key barrier of adoption.[42] Although strides are being made to address such concerns,[43] accountability still remains with the point of data collection, therefore requiring visibility and transparency as central requirements of a data environment.

Not merely limited to cloud-based data warehouses, visibility and transparency likewise must be observed throughout the lifecycle of data. Common to discussions on visibility and transparency, it is important to understand where the data is coming from, where it is going, how (and by whom) it will be used, and how it will be returned or destroyed. When any of these questions cannot be answered definitively, there resides some degree of privacy risk. While visibility and transparency are central to the foundational principles of PbD, there should not be confusion around

---

[39] *Id.*

[40] *Id.*

[41] *Id.*

[42] Andrew Charlesworth, *Accountability as a Way Forward for Privacy Protection in the Cloud*, *in* CLOUD COMPUTING 131, 144 (M.G. Jaatun et al. eds., 2009).

[43] Vishal R. Pancholi & Dr. Bhadresh P. Patel, *Enhancement of Cloud Computing Security with Secure Data Storage*, 2 INT'L J. INNOVATIVE RES. SCI. & TECH. 1, 1 (2016).

the necessity to encrypt, obscure, or otherwise properly protect data.[44] Whether traditional PII data, user data, or other potentially sensitive data, there is an inherent duty to safeguard the privacy interest in the data collected, stored, or used.[45]

G. Respect for User Privacy—Keep it User-Centric

Another notable aspect about the aforementioned Facebook response[46] is the company's focus on user-centric privacy protection. In many use cases, a company, university, or agency must consider not only end user access administration and privileges, but also the privacy interest inherent in the user data at hand.[47] In other words, it is a multistep or tiered approach that is often necessary to segregate end users from accessing data when building privacy conscious data-rich environments or solutions.

Risk emerges in data collection, storage, or use when there is an assumption that the privacy interest has detached because data has been de-identified.[48] In such cases, de-identification may be a mitigating factor, but it is not the sole factor of consideration.[49] Keeping a user-centric PbD approach also may mean employing strict data minimization strategies and maintaining a watchful eye for unintended biases to emerge in data use.

By combining each of these foundational principles into a cohesive PbD strategy, businesses, universities, and agencies can comprehensively address the privacy interests inherent in their Big Data environments.

H. Big Data and Regulatory Enforcement of the Privacy Interest in Data

FTC v. Wyndham demonstrates the doctrine of equity emerging as a leading factor in privacy-based regulatory enforcement actions.[50] Most recently, in the Federal Trade Commission (FTC) report, *Big Data: A Tool for Inclusion or Exclusion?*,[51] the FTC gave new color to the various laws, including the Fair Credit Reporting Act (FCRA),[52] equal opportunity laws,[53] and the Federal Trade Commission Act[54] as applicable to big data practices. Further, the report asserts,

---

[44] Cavoukian, *supra* note 7, at 4.

[45] *Id.*

[46] Protalinski, *supra* note 38.

[47] *Data De-Identification*, *supra* note 22, at 2-3.

[48] *Id.*

[49] *Id.*

[50] FTC v. Wyndham Worldwide Corp., 799 F.3d 236, 245 (3d Cir. 2015).

[51] FTC, BIG DATA: A TOOL FOR INCLUSION OR EXCLUSION?: UNDERSTANDING THE ISSUES 1 (2016), https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf.

[52] *Id.* at 13-17.

[53] *Id.* at 17-21.

[54] *Id.* at 21-23.

> Companies engaging in big data analytics should consider whether they are violating any material promises to consumers—whether that promise is to refrain from sharing data with third parties, to provide consumers with choices about sharing, or to safeguard consumers' personal information—or whether they have failed to disclose material information to consumers. In addition, companies that maintain big data on consumers should take care to reasonably secure consumers' data. Further, at a minimum, companies must not sell their big data analytics products to customers if they know or have reason to know that those customers will use the products for fraudulent or discriminatory purposes. The inquiry will be fact-specific, and in every case, the test will be whether the company is offering or using big data analytics in a deceptive or unfair way.[55]

The report additionally acknowledges the lifecycle of data, stating that, "The life cycle of big data can be divided into four phases: (1) collection; (2) compilation and consolidation; (3) data mining and analytics; and (4) use."[56]

Here, as a paramount point in consideration of PbD, the FTC acknowledges that not all data starts as Big Data, which is to suggest that the source of Big Data may in fact be derived from several smaller data sources.[57] Namely the FTC notes,

> As consumers browse the web or shop online, companies can track and link their activities. Sometimes consumers log into services or identify themselves when they make a purchase. Other times, techniques such as tracking cookies, browser or device fingerprinting, and even history sniffing identify who consumers are, what they do, and where they go. In the mobile environment, companies track and link consumers' activities across applications as another method of gathering information about their habits and preferences. More broadly, cross-device tracking offers the ability to interact with the same consumer across her desktop, laptop, tablet, wearable, and smartphone, using both online and offline information. Companies also are gathering data about consumers across the Internet of Things—the millions of Internet-connected devices that are in the market. Finally, data collection occurs offline as well, for example, through loyalty programs, warranty cards, surveys, sweepstakes entries, and even credit card purchases.[58]

With such a broad array of potential source data, the privacy risks are abundant. Thus, it is no surprise that the FTC is not the only regulator focused on this emerging area.[59] It is important to note that the PbD approach is also rooted in international regulatory frameworks, including the Resolution on Privacy by Design passed by the

---

[55] *Id.* at iv.

[56] *Id.* at 3.

[57] *Id.*

[58] *Id.* at 3-4.

[59] Council Regulation 2016/679, art. 24, 2016 O.J. (L 119) 1, 47 (EU) [hereinafter Council Regulation].

3*2nd International Conference of Data Protection and Privacy Commissioners*[60] and in the European General Data Protection Regulation under 3.4.4.1. Section 1—General Obligations, Article 23.[61]

In the U.S., the FTC was the leading regulator to recognize PbD in its report, Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Business and Policymakers.[62] In this report, as a baseline principle, the FTC recommended that, "Companies should promote consumer privacy throughout their organizations and at every stage of the development of their products and services."[63] The report additionally recognized as a substantial principle that, "Companies should incorporate substantive privacy protections into their practices, such as data security, reasonable collection limits, sound retention and disposal practices, and data accuracy."[64] And, as Procedural Protections to Implement the Substantive Principles, the report recommended that, "Companies should maintain comprehensive data management procedures throughout the life cycle of their products and services."[65]

Although U.S. federal and state lawmakers have yet to enact laws that expressly address the use of Big Data and the emerging Internet of Things, officials from both the FTC and Consumer Financial Protection Bureau (CFPB) have publicly acknowledged that this deficiency is not a significant impediment to their ability to act against companies that inappropriately handle consumer data.[66] Here, it should also be noted that there is unsettled civil case law supporting the notion that Big Data aggregators have been considered Consumer Reporting Agencies, as was the case in *Robins v. Spokeo*.[67] In that case, the court held that the plaintiff's allegations that the defendant, Spokeo, "regularly accept[ed] money in exchange for reports that 'contain[ed] data and evaluations regarding consumers' economic wealth and creditworthiness" were sufficient to support a plausible inference that defendant's conduct fell within the scope of the FCRA so as to survive the defendant's motion to dismiss.[68]

---

[60] 32nd Int'l Conference of Data Prot. & Privacy Comm'rs, *Resolution on Privacy by Design* 1-2 (2010), https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Cooperation/Conference_int/10-10-27_Jerusalem_Resolutionon_PrivacybyDesign_EN.pdf.

[61] Council Regulation, *supra* note 59, at 47.

[62] FTC, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS vii (2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

[63] *Id.*

[64] *Id.*

[65] *Id.*

[66] Allison Grande, *FTC, CFPB Setting Sights on 'Big Data' Enforcement*, LAW360 (May 11, 2015), http://www.law360.com/articles/654132/ftc-cfpb-setting-sights-on-big-data-enforcement.

[67] Civil Mins. 4, Robins v. Spokeo, Inc., No. CV10-05306 ODW(AGRX), 2011 U.S. Dist. LEXIS 14079, 2011 WL 597867 (C.D. Cal. Jan. 27, 2011).

[68] *Id.*

Alternatively, *Sweet v. LinkedIn* carved out a notable FCRA distinction based on the fact that the plaintiff user, Sweet, willingly provided her private information to the Big Data environment of LinkedIn.[69] In *Sweet*, the plaintiff's complaint did not establish the inference that LinkedIn gathers the information about the employment histories of the subjects of the Reference Searches (a paid LinkedIn service) to make consumer reports.[70] Rather, the court found that the intent was to carry out consumers' information-sharing objectives; therefore, the court did not find a plausible inference that LinkedIn acted as a consumer reporting agency with regard to its assembly of information.[71]

While the FCRA is but one U.S. law courts have interpreted in the dawn of Big Data, there are a number of U.S. federal laws and regulations more broadly focused on various aspects of privacy, which include, in part: Title III of the Omnibus Crime Control and Safe Streets Act of 1968,[72] Family Educational Rights and Privacy Act of 1974,[73] Health Insurance Portability and Accountability Act of 1996,[74] Privacy Act of 1974,[75] Right to Financial Privacy Act of 1978,[76] Privacy Protection Act of 1980,[77] Cable Communications Policy Act of 1984,[78] Electronic Communications Privacy Act of 1986,[79] Computer Matching and Privacy Protection Act of 1988,[80] Employee Polygraph Protection Act of 1988,[81] Video Privacy Protection Act of 1988,[82] Telemarketing Protection Act of 1991,[83] and the Gramm-Leach-Bliley Act, also known as the Financial Services Modernization Act of 1999.[84] In addition to U.S. privacy regulations, many states have adopted privacy-focused laws and regulations.[85]

---

[69] Sweet v. LinkedIn Corp., No. 5:14-CV-04531-PSG, 2015 U.S. Dist. LEXIS 49767, 2015 WL 1744254, at *6 (N.D. Cal. Apr. 14, 2015).

[70] *Id.*

[71] *Id.*

[72] 18 U.S.C. § 2510-22 (2016).

[73] 20 U.S.C. § 1232g (2016).

[74] 45 C.F.R. § 164.502 (2016).

[75] 5 U.S.C. § 552a (2016).

[76] 12 U.S.C. § 3402 (2016).

[77] 42 U.S.C. § 2000aa-6 (2016).

[78] 47 U.S.C. § 551 (2016).

[79] 18 U.S.C. § 2511 (2016).

[80] 5 U.S.C. § 552a (2016).

[81] 29 U.S.C. § 2002 (2016).

[82] 18 U.S.C. § 2710 (2016).

[83] 47 C.F.R. § 64.1200 (2016).

[84] 15 U.S.C. § 6801 (2016).

[85] Timothy A. Hartin, *Balancing Federal and Wisconsin Medical Privacy Laws*, 76 WIS. LAWYER 10, 50 (2003); *see also* Susan P. Stuart, *A Local Distinction: State Education Privacy Laws for Public Schoolchildren*, 108 W. VA. L. REV. 361, 380 (2005).

The State of Florida, while not expressly adopting a PbD approach, has adopted a privacy law that gives color to PII and establishes a framework for redress.[86] Under Florida Statute 501.171,

> 'Personal information' means either of the following: An individual's first name or first initial and last name in combination with any one or more of the following data elements for that individual: A social security number; A driver license or identification card number, passport number, military identification number, or other similar number issued on a government document used to verify identity; A financial account number or credit or debit card number, in combination with any required security code, access code, or password that is necessary to permit access to an individual's financial account; Any information regarding an individual's medical history, mental or physical condition, or medical treatment or diagnosis by a health care professional; or An individual's health insurance policy number or subscriber identification number and any unique identifier used by a health insurer to identify the individual.[87]

Additionally, Florida's statute includes a "user name or e-mail address, in combination with a password or security question and answer that would permit access to an online account."[88]

Uniquely, the Florida Statute, while not taking a PbD approach, does reward companies, universities, and agencies by including key exceptions for those that have implemented data protection strategies.[89] For example, the statute notes that,

> The term [PII] does not include information about an individual that has been made publicly available by a federal, state, or local governmental entity. The term also does not include information that is encrypted, secured, or modified by any other method or technology that removes elements that personally identify an individual or that otherwise renders the information unusable.[90]

Therefore, in the U.S., what is emerging in the absence of a consolidated express federal statute protecting the privacy interests of citizens in today's emerging Big Data era is an interwoven patchwork of privacy laws and regulations that lack clarity and may introduce greater litigation risk due to uncertainty.[91]

I. Global Privacy Compliance in the Big Data Era

Although this paper has given much attention to the U.S. privacy framework, the same lack of clarity and uncertainties are multiplied when assessed on a global scale.

---

[86] FLA .STAT. § 501.171 (2014).

[87] *Id.*

[88] *Id.*

[89] *Id.*

[90] *Id.*

[91] David J. Walton, *Big Data Raises Big Legal Issues*, INSIDE COUNSEL (Mar. 28, 2014), http://www.insidecounsel.com/2014/03/28/big-data-raises-big-legal-issues

For companies, universities, and agencies that are faced with the realities of the cross-border movement of PII data, proactively adopting a PbD approach can be the best investment to mitigate privacy risk and to ensure global privacy compliance.

As a solution in global privacy compliance, PbD is an evolving framework with applicability to the continual advancement of data collection, storage, and use. PbD is a relatively recent construct, yet its prevalence in the continued development of data-rich environments is imminent as a measure of addressing the lack of clarity and uncertainty that exist in the current state of privacy law.[92] When implementing a PbD approach, strong global privacy compliance programs are aptly defined by strong governance, transparency, and reporting.[93]

Dealing with matters of global privacy can be very challenging, giving rise to third party specialization.[94] Independent third party companies like PwC recognize the value of a PbD framework in noting that "with global privacy compliance, the inquiry needs to focus on the entire data lifecycle, from collection through transmission, access, storage, use and destruction."[95] As PwC highlights,[96] key questions they focus on in global privacy assessments include:

> • Is there a national, over-arching privacy or data protection law in a certain country, does it apply to the client and its data, are there any industry-specific requirements, and what does the client need to do to demonstrate compliance?
> • Are the foreign regulatory requirements readily available (particularly in English) to permit a prompt and effective risk assessment?
> • Are "personal data" and "sensitive data" or similar terms defined in the country's requirements, and what data elements are considered "personal" or "sensitive" in specific countries, thereby potentially requiring clients to implement heightened safeguards?
> • Must any specific notice be given to data subjects about the purposes for which data is collected, how data is used, or the company's privacy practices?
> • What rights of access do data subjects have to their data?
> • What rights of correction/redress do data subjects have with regard to inaccurate or incomplete data?
> • Can customer, employee, or business partner data collected in a certain country be transmitted out of that country, processed in another country, and freely flow back home again?
> • What specific privacy or data security safeguards, if any, are required?
> • Is there a data protection authority with which registration is required?

---

[92] *Have it all: Protecting Privacy in the Age of Analytics*, DELOITTE, http://www2.deloitte.com/content/dam/Deloitte/ca/Documents/Analytics/ca-en-analytics-ipc-big-data.pdf (last visited Oct. 2, 2016) [hereinafter *Have it all*].

[93] *See id.*

[94] PwC Advisory, *Key Considerations in Financial Services Global Privacy Compliance*, PwC (2007), http://www.pwc.com/us/en/banking-capital-markets/publications/assets/global_privacy_compliance.pdf.

[95] *Id.*

[96] *Id.*

> • Can a regulator or other government official demand access to certain data, search data processing facilities, or stop the client from using or transmitting certain data?
> • Are notifications required in the event of a data security breach?
> • Is privacy-related training required for personnel?
> • Can vendors outsource certain processes involving personal data, and are there specific due diligence, contractual, or oversight requirements?
> • What risks or sanctions does the client face if it fails to comply with a country's privacy or data protection laws?[97]

Here, again, we can see the value of the PbD foundational principles in addressing these unique privacy challenges. PwC also notes,[98] fundamentally, the conversation comes down to the data itself:

> •**Whose** personal or sensitive data do we collect, store, transmit and use?
> •**What** specific data elements are located in certain applications or systems?
> •**When** is personal or sensitive data transmitted across national borders?
> •**Where** in various business processes is personal or sensitive data used?
> •**Why** are we sharing personal or sensitive data with affiliates or third-parties?
> •**How** is personal or sensitive data transmitted (encrypted?) and stored at rest (encrypted?)?[99]

As we address these foundational questions within the PbD framework, we become able to build robust and meaningful global privacy programs that take into account the holistic lifecycle of the Big Data environments and data uses.

As identified, the sources of Big Data are diverse and are continually emerging. Thus, in the broader context of global privacy, embracing a PbD approach can be the best investment to mitigate privacy risk and to ensure global privacy.

J. The Right Thing to Do

Why should businesses, universities, or agencies adopt a PbD approach? The answer to this question actually has little to do with the legal or regulatory frameworks. It is much simpler than that; PbD is about establishing trust and respect, and its adoption is simply the right thing to do.

As a case in point, in today's Big Data era, Apple has emerged among the most consumer-trusted companies in technology because the company has adopted a PbD approach.[100] As Apple confirms, "Security and privacy are fundamental to the design of all our hardware, software, and services, including iCloud and new services like Apple Pay."[101] Apple also says, "At Apple, your trust means everything to us. That's why we respect your privacy and protect it with strong

---

[97] *Id.*

[98] *Id.*

[99] *Id.* (Emphasis in original).

[100] *Privacy*, APPLE.COM, http://www.apple.com/privacy/ (last visited Aug. 24, 2016).

[101] *Id.*

encryption, plus strict policies that govern how all data is handled."[102] Demonstrating their user-centric PbD approach, the company notes, "We believe in telling you up front exactly what's going to happen to your personal information and asking for your permission before you share it with us."[103]

In the Apple approach to privacy, PbD has been so engrained in the operation of the business that it has emerged as a competitive differentiator for the Apple brand.[104] When looking for the root of Apple's focus on PbD, what is notable is that it is not focused on the legal frameworks; rather, more simply, it notes, "Our commitment to protecting your privacy comes from a deep respect for our customers."[105] Apple's privacy approach method is not tied to the legal or regulatory frameworks;[106] alternatively, the Apple PbD approach always is focused on doing the right thing for its customers.[107] Moreover, its strong stance on user-centric privacy has increasingly put the company at odds with lawmakers and the law enforcement community.[108]

In May of 2015, signing a joint letter to the President of the United States, Apple partnered in declaring,

> More than undermining every American's cybersecurity and the nation's economic security, introducing new vulnerabilities to weaken encrypted products in the U.S. would also undermine human rights and information security around the globe. If American companies maintain the ability to unlock their customers' data and devices on request, governments other than the United States will demand the same access, and will also be emboldened to demand the same capability from their native companies. The U.S. government, having made the same demands, will have little room to object. The result will be an information environment riddled with vulnerabilities that could be exploited by even the most repressive or dangerous regimes. That's not a future that the American people or the people of the world deserve.[109]

PbD reminds us to consider the win-win approach to building privacy conscious, data-rich environments or solutions. As companies, universities, and agencies operating in today's Big Data environment define their own privacy strategies, PbD

---

[102] *Id.*

[103] *Id.*

[104] *Id.*

[105] *Id.*

[106] *Id.*

[107] *Id.*

[108] Julia Love, *Apple 'Privacy Czars' Grapple with Internal Conflicts over User Data,* REUTERS (Mar. 21, 2016), http://www.reuters.com/article/us-apple-encryption-privacy-insight-idUSKCN0WN0BO.

[109] Christian Dawson, *i2Coalition Signs Open Letter on Encryption and Privacy*, I2 INTERNET INFRASTRUCTURE COALITION (May 19, 2015), http://www.i2coalition.com/i2coalition-signs-open-letter-on-encryption-and-privacy/.

provides a framework for taking control of Big Data in how we collect, store, and use this valuable asset as it is simply the right thing to do.[110]

IV. CONCLUSION

If anything is certain in the modern Big Data environment, it is that what is considered Big Data today is but miniscule in contrast to future data environments. PbD provides a deep-rooted framework that can be broadly applied from governance policy through design. While also providing a strong toolset for maintaining global privacy, PbD is simply the right thing to do.

In review, the foundational principles of PbD include: 1) proactive not reactive, preventative not remedial; 2) privacy as the default setting; 3) privacy embedded into design; 4) full functionality—positive-sum, not zero-sum; 5) end-to-end security—full lifecycle protection; 6) visibility and transparency—keep it open; and 7) respect for user privacy—keep it user-centric.[111] It is through the examination of each of these core tenants that the applicability of PbD to this emerging area of Big Data is best considered. For companies, universities, and agencies that are faced with the global realities of the cross border movement of PII data, proactively adopting a PbD approach can be the best investment to mitigate privacy risk and to ensure global privacy interests are addressed. As a solution, PbD is an evolving framework with applicability to the continual advance of data collection, storage, and use.

Although PbD is a recently emerging framework, it is deeply rooted, and the opportunity in the continued development of building privacy conscious, data-rich environments or solutions is imminent as a measure of addressing the lack of clarity and uncertainty that exist in the current state of privacy law. Among the most influential techniques of addressing PbD in the Big Data environment are de-identification and data minimization;[112] however, despite their effectiveness, these techniques are highly limited without the benefit of a robust PbD culture and supporting governance framework.[113] Looking at the competitive landscape of the technology industry, those companies that have successfully adopted a PbD culture are also those that have found competitive differentiation for their brands.[114] The focus for such companies is rarely the legal frameworks within which they operate; but instead, the underlying motivation is doing the right thing for their customers by respecting their data and privacy interests.[115]

The PbD framework provides an approach that allows businesses, universities, and agencies to focus on respecting the privacy interests of their users and allows them to consistently earn their trust by adopting security and privacy as fundamental to the design of all Big Data hardware, software, and services. The long arm of the law is continuing to stretch into Big Data, yet we can be certain that more legal and regulatory frameworks will emerge to better govern the increasing privacy risks inherent therein. Not only can we be certain that new legal frameworks will emerge,

---

[110] Cavoukian, *supra* note 7.

[111] *Id.*

[112] *Have it all*, *supra* note 93.

[113] *Id.*

[114] *Id.*

[115] *See, e.g.*, *Privacy*, *supra* note 101.

but we know that the sheer volume and velocity of future Big Data environments will continue to expand.[116] In conclusion, as we begin making decisions that will shape the future of the Big Data landscape, PbD offers a framework for establishing a privacy-first culture and governance approach that will not only address the uncertainties we face, but also the framework is simply the right thing to do.

[116] *Have it all*, *supra* note 93.