BIG DATA ETHICS

Neil M. Richards[*]
Jonathan H. King[**]

INTRODUCTION

We are on the cusp of a "Big Data" Revolution. Increasingly large datasets are being mined for important predictions and often surprising insights. We are witnessing merely the latest stage of the Information Revolution that has transformed our society and our lives over the past half century. But the big data phase of the revolution promises (or threatens, depending on one's perspective) a greater scale of social change at an even greater speed. The scale of the Big Data Revolution is such that all kinds of human activities and decisions are beginning to be influenced by big data predictions, including dating, shopping, medicine, education, voting, law enforcement, terrorism prevention, and cybersecurity. This transformation is comparable to the Industrial Revolution in the ways our pre–big data society will be left radically changed.

The potential for social change means that we are now at a critical moment; big data uses today will be sticky and will settle both default norms and public notions of what is "no big deal" regarding big data predictions for years to come. Individuals have little idea concerning what data is being collected, let alone shared with third parties. Existing privacy protections focused on managing personally identifying information are not enough when secondary uses of big data sets can reverse engineer past, present, and even future breaches of privacy, confidentiality, and identity.[1] Many of the most revealing personal data sets such as call history, location history, social network connections, search history, purchase history, and facial recognition are already in the hands of governments and corporations. Further, the collection of these and other data sets is only accelerating.

---

* Professor of Law, Washington University. We would like to thank Ujjayini Bose, Matthew Cin, and Carolina Foglia for their very helpful research assistance.

** LLM Graduate in Intellectual Property and Technology Law, Washington University and Vice President of Cloud Strategy and Business Development for CenturyLink Technology Solutions. The views and opinions expressed by the author are not necessarily the views of his employer.

1. *See* Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1881 (2013).

As the amount and variety of data continue to grow, defining the catchall term "big data" can be elusive. Technical definitions of big data are often narrowly constrained to describe "data that exceeds the processing capacity of conventional database systems."[2] Technologists often use the technical "3-V" definition of big data as "high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."[3] Peter Mell, a computer scientist with the National Institute of Standards and Technology, similarly constrains big data to "[w]here the data volume, acquisition velocity, or data representation limits the ability to perform effective analysis using traditional relational approaches or requires the use of significant horizontal scaling for efficient processing."[4]

We prefer to define big data and big data analytics socially, rather than technically, in terms of the broader societal impact they will have. Mayer-Schönberger and Cukier define big data as referring "to things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value, in ways that change markets, organizations, the relationship between citizens and governments, and more."[5] We have some reservations about using the term "big data" at all, as it can exclude important parts of the problem, such as decisions made on small data sets, or focus us on the size of the data set rather than the importance of decisions made based upon inferences from data. Perhaps "data analytics" or "data science" are better terms, but in this paper we will use the term "big data" (to denote the collection and storage of large data sets) and "big data analytics" (to denote inferences and predictions made from large data sets) consistent with what we understand the emerging usage to be.

In a prior article, we argued that nontransparent collection of small data inputs enables big data analytics to identify, at the

---

2. Edd Dumbill, *What Is Big Data?: An Introduction to the Big Data Landscape*, O'REILLY (Jan. 11, 2012), http://strata.oreilly.com/2012/01/what-is-big-data.html.

3. *IT Glossary: Big Data*, GARTNER, http://www.gartner.com/it-glossary/big-data/ (last visited Feb. 23, 2014). For the original "3-Vs" Gartner report, see Doug Laney, *3D Data Management: Controlling Data Volume, Velocity, and Variety*, GARTNER (Feb. 6, 2001), http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf. Gartner has also classified big data at the peak of its "Hype Cycle." *See* Arik Hesseldahl, *Think Big Data Is All Hype? You're Not Alone*, ALL THINGS D (Aug. 19, 2013, 11:54 AM), http://allthingsd.com/20130819/think-big-data-is-all-hype-youre-not-alone/.

4. Frank Konkel, *Sketching the Big Picture on Big Data*, FCW (Apr. 15, 2013), http://fcw.com/articles/2013/04/15/big-experts-on-big-data.aspx?m=1.

5. *See* VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK, AND THINK 6 (2013).

expense of individual identity, and empower institutions that possess big data capabilities.[6] In this paper, we argue that big data, broadly defined, Is producing increased powers of institutional awareness and power that require the development of Big Data Ethics. We are building a new digital society, and the values we build or fail to build into our new digital structures will define us. Critically, if we fail to balance the human values that we care about, like privacy, confidentiality, transparency, identity, and free choice, with the compelling uses of big data, our big data society risks abandoning these values for the sake of innovation and expediency.

Our argument proceeds in three Parts. In Part I, we trace the origins and rapid growth of the Information Revolution and describe how we as a society have effectively built a "big metadata computer" that is now computing data and associated metadata about everything we do at an ever quickening pace. As the data about everything (including us) have grown, so too have big data analytics—new capabilities enable new kinds of data analysis and motivate increased data collection and the sharing of data for secondary uses. Using examples taken from the Big Data Revolution, we show how government institutions are already adopting big data tools to strengthen their awareness about (and by extension their power over) the world.

In Part II, we call for the development of "Big Data Ethics," a set of four high-level principles that we should recognize as governing data flows in our information society, and which should inform the establishment of legal and ethical big data norms. To advance ethics of big data, four such principles should be paramount.

First, *we must recognize "privacy" as information rules.* We argue that privacy in the age of big data should be better understood as the need to expand the rules we use to govern the flows of personal information. We show how the prophesy that "privacy is dead" is misguided. Even in an age of surveillance and big data, privacy is neither dead nor dying. Notions of privacy are changing with society as they always have. But privacy (and privacy law) are very much alive; while the amount of personal information that is being recorded is certainly increasing, so too is the need for rules to govern this social transformation. Understanding privacy rules as merely the ability to keep information secret severely handicaps our ability to comprehend and shape our digital revolution. What has failed is not privacy but what Daniel Solove has termed "Privacy Self-Management," the idea that it is possible or desirable for every individual to monitor and manage a shifting collection of privacy

6. Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 42–43 (2013).

settings of which they may only be dimly aware.[7] We argue that "privacy" in today's information economy should be better understood as encompassing information rules that manage the appropriate flows of information in ethical ways.

Second, *we must recognize that shared private information can remain "confidential."* Much of the tension in privacy law over the past few decades has come from the simplistic idea that privacy is a binary, on-or-off state, and that once information is shared and consent given, it can no longer be private. Binary notions of privacy are particularly dangerous and can erode trust in our era of big data and metadata, in which private information is necessarily shared by design in order to be useful. The law has always protected private information in intermediate states, whether through confidentiality rules like the duties lawyers and doctors owe to clients and patients; evidentiary rules like the ones protecting marital communications; or statutory rules like the federal laws protecting health, financial, communications, and intellectual privacies. Neither shared private data (nor metadata) should forfeit their ability to be protected merely because they are held in intermediate states. Understanding that shared private information can remain confidential better helps us see how to align our expectations of privacy with the rapidly growing secondary uses of big data analytics.

Third, *we must recognize that big data requires transparency.* Transparency has long been a cornerstone of civil society as it enables informed decision making by governments, institutions, and individuals alike. The many secondary uses of big data analytics, and the resulting incentives of companies and governments to share data, place heightened importance on transparency in our age of big data. Transparency can help prevent abuses of institutional power while also encouraging individuals to feel safe in sharing more relevant data to make better big data predictions for our society.

Fourth, *we must recognize that big data can compromise identity.* "Identity," like privacy, can be hard to define. We use identity to refer to the ability of individuals to define who they are. Big data predictions and inferences risk compromising identity by allowing institutional surveillance to identify, categorize, modulate, and even determine who we are before we make up our own minds. We must therefore begin to think imaginatively about the kinds of data inferences and data decisions we will allow. We must regulate or prohibit ones we find corrosive, threatening, or offensive to citizens, consumers, or individual humans, just as we have long protected decisions like voting and contraception and prohibited invidious decisions made upon criteria like race, sex, or gender.

How should we integrate Big Data Ethics into our society? In Part III, we suggest how this should be done. Law will be an

7. Solove, *supra* note 1, at 1880–81.

important part of Big Data Ethics, but so too must the establishment of ethical principles and best practices that guide government agencies, corporate actors, data brokers, information professionals, and individual humans, whether we label them "Chief Privacy Officer," "Civil Liberties Engineer," "system administrator," "employee," or "user." Individuals certainly share responsibility for ethical data usage and development, but the failure of the privacy-self-management system shows that we must build structures that encourage ethical data usage rather than merely nudging individual consumers into sharing as much as possible for as little as possible in return. Big Data Ethics are as much a state of mind as a set of mandates. While engineers in particular must embrace the idea of Big Data Ethics, in an information society that cares about privacy, we must all be part of the conversation and part of the solution.

I. THE BIG DATA REVOLUTION

The Big Data Revolution is the latest stage in the wider Information Revolution that is rapidly changing life around us. Building upon discoveries made during and after the Second World War, the Information Revolution rapidly picked up speed in the 1970s with Intel's invention of the microprocessor. If the first act of the Information Revolution was defined by the microprocessor and the power to compute, and the second by the network and the power to connect, the third will be defined by data and the power to predict. One way to look at things is that we have collectively built and are now living with a really big metadata computer.

A. The Big Metadata Computer

We have always been surrounded by information. We have also long had math and human "computers" to help us process and make sense of information. After World War II, however, urgent problems like nuclear weapon air defense spurred investment into new kinds of computers. These computers used innovations in communications and material sciences that enabled machine computers with transistors to reliably transfer, store, and retrieve information as data.[8] Uses for these early computers quickly expanded beyond military applications to meet insatiable corporate demand.

Early pioneers saw the human possibilities as well. In a famous 1950 article, Alan Turing suggested that one day computer processing might become so powerful as to be externally indistinguishable from human thought.[9] J.C.R. Licklider predicted in a 1960 paper entitled *Man-Computer Symbiosis* that "in not too

8. M. MITCHELL WALDROP, THE DREAM MACHINE: J.C.R. LICKLIDER AND THE REVOLUTION THAT MADE COMPUTING PERSONAL 113 (2001).

9. *See* A.M. Turing, *Computing Machinery and Intelligence*, 59 MIND 433, 460 (1950).

many years, human brains and computing machines will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today."[10] Licklider optimistically believed that man-computer symbiosis would be "intellectually the most creative and exciting in the history of mankind."[11]

Gordon Moore, then head of research and development for Fairchild Semiconductor, observed in a 1965 article that the number of transistors on a chip had roughly doubled each year from 1959 to 1965.[12] Moore grasped the mathematical significance of such exponential progress and predicted that this phenomenon would enable "such wonders as home computers—or at least terminals connected to a central computer—automatic controls for automobiles, and personal portable communications equipment."[13] Moore's article also first articulated what is now referred to as "Moore's Law," the prediction that the number of transistors on a chip would roughly double every two years.[14]

Processors doubling in computing power every two years also came with a corresponding decrease in the cost of computing. Lower costs of computing led to the development of ever more powerful software taking advantage of ever more powerful hardware. Half a century on, Moore's law and others like it have enabled the migration of computing from its military and corporate roots into the hands of virtually everyone in the developed world. Bill Gates's ambitious 1980s vision of "a computer on every desk and in every home" has already come and gone.[15] We have moved on to the smartphone and tablet era, ushered in by Apple's triumphant transformation from a computer company into "a mobile device company."[16]

---

10. J.C.R. Licklider, *Man-Computer Symbiosis*, HFE-1 IRE TRANSACTIONS ON HUM. FACTORS ELECTRONICS 4, 4 (1960), *available at* http://worrydream.com/refs/Licklider%20-%20Man-Computer%20Symbiosis.pdf.

11. *Id.* at 5.

12. Gordon E. Moore, *Cramming More Components onto Integrated Circuits*, 38 ELECTRONICS 114, 114 (Apr. 19, 1965), *available at* http://web.eng.fiu.edu/npala/EEE6397ex/Gordon_Moore_1965_Article.pdf.

13. *Id.* at 114.

14. BILL GATES, THE ROAD AHEAD 31 (1995); Jon Stokes, *Classic.Ars: Understanding Moore's Law*, ARS TECHNICA (Sept. 27, 2008, 9:00 AM), http://arstechnica.com/gadgets/2008/09/moore/.

15. Claudine Beaumont, *Bill Gates's Dream: A Computer in Every Home*, TELEGRAPH (June 27, 2008, 12:01 AM), http://www.telegraph.co.uk/technology/3357701/Bill-Gatess-dream-A-computer-in-every-home.html.

16. Steve Jobs, Speech Given at the Unveiling of the New Apple iPad (Jan. 2010), *available at* http://www.apple.com/apple-events/january-2010/ (noting that "Apple is the largest mobile devices company in the world now"); *see also* Erick Schonfeld, *Tim Cook: Apple is "A Mobile-Device Company,"* TECHCRUNCH

Now, at breakneck pace, computing is distributing to everything and "software is eating the world."[17] Governments and corporations are rapidly adopting Infrastructure as a Service ("IaaS"), also referred to as cloud computing. Even NASA uses cloud computing to help it conduct missions to land rovers on Mars.[18] New digital delivery businesses either embrace the cloud, like the former mail-order business Netflix has done, or they fail to adapt and, like Blockbuster, go out of business.[19] Personal computing power is moving into smartphones, tablets, and wearable devices.[20] A "Quantified Self" movement allows people to measure their lives to help improve sleep and lose weight. The machines we use, the new things we buy, and, it seems, "everything" increasingly holds increasing amounts of computational power.[21]

This computational power is also fueling unprecedented growth in applications and software tools of all kinds. Since launching in July 2008, the Apple App Store has grown to an inventory of close to one million applications ("apps"), with tens of thousands of new apps added every month.[22] Apple's App Store ranking algorithms constantly adjust to keep up.[23] Overtaking Apple's head start, the Google Play store for Android already crossed the million app milestone in July 2013.[24] Leveraging the on-demand scale and power of cloud computing, an entire new model of software delivery has also emerged called Software as a Service ("SaaS"), which one

---

(Feb. 23, 2010), http://techcrunch.com/2010/02/23/tim-cook-apple-mobile-device-company/.

17. Marc Andreessen, *Why Software Is Eating the World,* WALL ST. J., Aug. 20, 2011, at C2.

18. Andrea Chang, *NASA Uses Amazon's Cloud Computing in Mars Landing Mission*, L.A. TIMES (Aug. 9, 2012), http://articles.latimes.com/2012/aug/09/business/la-fi-tn-amazon-nasa-mars-20120808.

19. Ben Mauk, *Last Blues for Blockbuster*, NEW YORKER (Nov. 8, 2013), http://www.newyorker.com/online/blogs/currency/2013/11/remembering-blockbuster-with-little-nostalgia.html.

20. Bill Wasik, *Why Wearable Tech Will Be as Big as the Smartphone*, WIRED (Dec. 17, 2013, 6:30 AM), http://www.wired.com/gadgetlab/2013/12/wearable-computers/.

21. *See generally* Dave Evans, *The Internet of Everything: How More Relevant and Valuable Connections Will Change the* World, CISCO (2012), http://www.cisco.com/web/about/ac79/docs/innov/IoE.pdf.

22. Chuck Jones, *Apple's App Store About to Hit 1 Million Apps*, FORBES (Dec. 11, 2013, 12:53 PM), http://www.forbes.com/sites/chuckjones/2013/12/11/apples-app-store-about-to-hit-1-million-apps/.

23. Sarah Perez, *Widespread Apple App Store Search Rankings Change Sees iOS Apps Moved over 40 Spots, on Average*, TECHCRUNCH (Dec. 13, 2013), http://techcrunch.com/2013/12/13/widespread-apple-app-store-search-rankings-change-sees-ios-apps-moved-over-40-spots-on-average/.

24. Christina Warren, *Google Play Hits 1 Million* Apps, MASHABLE (July 24, 2013), http://mashable.com/2013/07/24/google-play-1-million/.

leading industry analyst predicts will grow to $75 billion in 2014.[25] Right behind SaaS, developers now rapidly create custom-built applications on Platform as a Service ("PaaS") offerings.

Connecting this staggering amount of distributed computing, running ever-multiplying numbers of applications, is an equally astonishing global communications network. The Internet also outpaced its military origins and quickly spread to connect academia, corporations, individuals, and now physical devices in our cities and homes. Cisco reports that global Internet Protocol ("IP") traffic has increased fourfold in the last five years and that there will be nearly three times as many devices connecting to IP networks as the global population by 2017.[26] In November 2013, Ericsson reported total mobile subscriptions of 6.6 billion and 40% growth in the number of these subscriptions annually.[27] Keeping up with these connecting devices, we have depleted the 4.2 billion unique IP addresses in IP version four, requiring us to switch to IP version six, with a potential three hundred and forty trillion addresses.[28]

From telegraph to the Internet,[29] global communications now surge through over 550,000 miles of undersea fiber-optic cables.[30] From telecommunications provider to content provider, players like Google, Facebook, Microsoft, and Amazon are now building their own fiber-optic networks to have more control over their content and their economics.[31] In the air around us, what was once wireless spectrum for UHF TV is now "beachfront" spectrum being auctioned for billions of dollars because it can more easily penetrate buildings to enhance connectivity and communication.[32] In the air above us,

---

25. Alex Williams, *Forrester: SaaS and Data-Driven "Smart" Apps Fueling Worldwide Software Growth*, TECHCRUNCH (Jan. 3, 2013), http://techcrunch.com/2013/01/03/forrester-saas-and-data-driven-smart-apps-fueling-worldwide-software-growth/.

26. *Cisco Visual Networking Index: Forecast and Methodology, 2012–2017*, CISCO 1 (May 29, 2013), http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf.

27. *See Ericsson Mobility Report: On the Pulse of Networked Society*, ERICSSON 4 (Nov. 2013), http://www.ericsson.com/res/docs/2013/ericsson-mobility-report-november-2013.pdf.

28. *World Tests IPv6: Why 4.2 Billion Internet Addresses Just Weren't Enough* (June 8, 2011), *available at* http://www.pbs.org/newshour/bb/science/jan-june11/ipv6_06-08.html.

29. *See generally* TOM STANDAGE, THE VICTORIAN INTERNET: THE REMARKABLE STORY OF THE TELEGRAPH AND THE NINETEENTH CENTURY'S ON-LINE PIONEERS (1998).

30. Todd Lindeman, *A Connected World*, WASH. POST (July 6, 2013), http://apps.washingtonpost.com/g/page/business/a-connected-world/305/.

31. Drew FitzGerald & Spencer E. Ante, *Tech Firms Push to Control Web's Pipes*, WALL ST. J. (Dec. 16, 2013, 8:36 PM), http://online.wsj.com/news/articles/SB10001424052702304173704579262361885883936.

32. Philip J. Weiser & Dale Hatfield, *Spectrum Policy Reform and the Next Frontier of Property Rights*, 15 GEO. MASON L. REV. 549, 549, 578 (2008).

over 1,000 satellites operate.[33] The United States Air Force ensures that twenty-four of these satellites provide GPS signals so our mobile devices can almost always know where in the world they are located.[34] Self-service Wi-Fi has grown astronomically. Think how quickly we all have been acculturated into asking, upon entering a room, "What's your Wi-Fi password?"

What are all these computers primarily computing and networks now primarily networking? Data, and lots of them. An often-cited standard unit of large amounts of data is the aggregate amount of information stored in the books of the Library of Congress.[35] In 1997, Michael Lesk, in his report "How Much Information Is There in the World," estimated that there were twenty terabytes of book data stored in the Library of Congress.[36] According to one of the documents leaked by Edward Snowden, the NSA was ingesting "one Library of Congress every 14.4 seconds" as early as 2006.[37]

Now the Library of Congress itself is collecting data, with 525 terabytes already in its web archive as of May 2014.[38] Twitter and the Library of Congress reached an agreement in April 2010 that enabled the library to archive public tweets since 2006.[39] As of January 2013, the Library of Congress had archived 130 terabytes, comprised of over 170 billion tweets and growing by nearly half a billion more tweets each day.[40]

The Library of Congress example reveals the growth not merely of data but of an important kind of data called "metadata." The Library is not merely collecting the 140 characters in each tweet. In

---

33. Fraser Cain, *How Many Satellites Are in Space?*, UNIVERSE TODAY (Oct. 24, 2013), http://www.universetoday.com/42198/how-many-satellites-in-space/.

34. *See* Mark Sullivan, *A Brief History of GPS*, TECHHIVE (Aug. 9, 2012, 7:00 AM), http://www.techhive.com/article/2000276/a-brief-history-of-gps.html (outlining a timeline of the use of GPS).

35. *See* Leslie Johnston, *How Many Libraries of Congress Does It Take?*, SIGNAL: DIGITAL PRESERVATION (Mar. 23, 2012), http://blogs.loc.gov/digitalpreservation/2012/03/how-many-libraries-of-congress-does-it-take/ (listing examples of references to the size of the Library of Congress).

36. MICHAEL LESK, HOW MUCH INFORMATION IS THERE IN THE WORLD? (1997), *available at* http://www.lesk.com/mlesk/ksg97/ksg.html.

37. Barton Gellman, *Edward Snowden: "I Already Won,"* WASH. POST, Dec. 24, 2013, at A1.

38. Scott Maucione, *Can Digital Data Last Forever?*, FEDSCOOP (Nov. 8, 2013, 8:00 AM), http://fedscoop.com/can-digital-data-last-forever/; *Web Archiving FAQs*, LIBR. CONGRESS, http://www.loc.gov/webarchiving/faq.html#faqs_05 (last visited Feb. 25, 2014).

39. LIBRARY OF CONGRESS, UPDATE ON THE TWITTER ARCHIVE AT THE LIBRARY OF CONGRESS 1 (2013), *available at* http://www.loc.gov/today/pr/2013/files/twitter_report_2013jan.pdf.

40. Rex W. Huppke, *170 Billion Saved Tweets Make a Tower of Babble*, CHI. TRIB., Jan 8, 2013, at 2; Doug Gross, *Library of Congress Digs into 170 Billion Tweets*, CNN (Jan. 7, 2013, 12:18 PM), http://www.cnn.com/2013/01/07/tech/social-media/library-congress-twitter/.

addition to the 140 characters of text, each tweet also has over thirty-one documented metadata fields.[41] Metadata is commonly defined as a set of data that describes and gives information about other data.[42] Thus, each tweet's metadata also reveals the identity of its author as well as the date, time, and location from which it was sent, among other things. This is metadata—data about data themselves.

We have of course long created metadata, such as the old card cataloging systems that libraries maintained for centuries. The creation (let alone storage) of metadata, however, usually required much effort and cost.[43] Librarians went through the laborious task of creating book metadata for library catalogs so that books could be more easily organized, found, and referenced. To allow the post office to deliver our mail, we take the time to write the recipient and return address metadata on our envelopes. When we started to speak by phone, the phone companies developed technology to record the metadata of the phone numbers we dialed, when the calls took place, and how long they lasted so they could place the call and properly bill us. Metadata makes phone calls possible. The time and effort to create metadata was worth it because it considerably increased the value of associated data (the book or the phone number) by allowing more opportunity for their use.

Today we live in a radically different metadata world. The combination of ever more powerful computing, networking, and data storage has enabled the automated and largely costless generation and collection of metadata with nearly everything we do. The envelopes we used to address are eclipsed by the e-mails we send. The analog phone calls we used to make have long since been converted to digital technologies, enabling inherent metadata creation and easier sharing as revealed by the NSA metadata collection programs.[44] Knowingly or unknowingly, with every Google search, every Facebook post, and even every time we simply turn on our smartphones (or move with them on), we produce metadata. Moreover, metadata about us are added to commercial algorithms like Facebook's Tag Suggest facial-recognition system to

41. *See* Paul Ford, *What Twitter's Made of*, BLOOMBERG BUSINESSWEEK, Nov. 11, 2013, at 12–13 (discussing the large amount of data that comes with a 140 character Tweet).

42. *See Metadata Definition*, DICTIONARY.COM, http:///dictionary.reference.com/browse/metadata?s=t (last visted May 5, 2014).

43. CATHERINE C. MARSHALL, MAKING METADATA: A STUDY OF METADATA CREATION FOR A MIXED PHYSICAL-DIGITAL COLLECTION (1998), *available at* http://www.csdl.tamu.edu/~marshall/dl98-making-metadata.pdf ("As surely as metadata is valuable, it is also difficult and costly to create.").

44. *See* Glenn Greenwald, *US Orders Phone Firm to Hand over Data on Millions of Calls*, GUARDIAN (Regional), June 6, 2013, at 1 (explaining a National Security Agency program which collects telephone records of Verizon customers).

make them more powerful.[45] Alessandro Acquisti has explained how Facebook and other publicly available sources of facial data combined with ubiquitous cloud computing and rapidly improving facial recognition capabilities will result in "a radical change in our very notions of privacy and anonymity."[46]

Stepping back, all of this distributed computing that is powering networked devices and applications generating Library of Congress multiples of data is starting to become a kind of big metadata computer. Individuals, companies, and governments collectively feed and interact with this big metadata computer every minute of every day. Further, rapidly improving hardware, software, protocols, and standards around this big metadata computer enable us to generate better metadata and share them more easily. We want to be clear here: we need and want this big metadata computer to thrive. Many of the marvels of the last few decades and of those to come depend upon its continued, rapid expansion. But like many new and powerful tools, the big metadata computer creates challenges; specifically, it allows new inferences, insights, and predictions that will create problems of their own.

B. Big Data Adoption

In the early days of data analysis, companies had to perform the time-intensive task of feeding internally generated data into data warehouses to improve data insights as "production processes, sales, customer interactions, and more were recorded, aggregated, and analyzed."[47] A new era of big data began when companies began to gather and analyze large amounts of information from internal and external sources. To meet the demands of storing and analyzing these larger data sets, innovators like Google, Yahoo, LinkedIn, and eBay developed new, open-source software technologies such as Hadoop, a software tool that allows the storage and processing of very large data sets across collections of computers.[48] Larger data sets enabled new possibilities of a radically different scale than in the past. Mayer-Schönberger and Cukier provide a helpful analogy here, stating, "[A] movie is fundamentally different from a frozen photograph. It's the same with big data: by changing the amount,

---

45. *See* Sophie Curtis, *Facebook Defends Using Profile Pictures for Facial Recognition*, TELEGRAPH (Nov. 15, 2013, 5:14 PM), http://www.telegraph.co.uk/technology/facebook/10452867/Facebook-defends-using-profile-pictures-for-facial-recognition.html.

46. Alessandro Acquisti, *Why Privacy Matters*, TED (June 2013), http://www.ted.com/talks/alessandro_acquisti_why_privacy_matters.html.

47. Thomas H. Davenport, *Analytics 3.0*, HARV. BUS. REV., Dec. 2013, at 66; *see also* Jeff Kelly, *Big Data: Hadoop, Business Analytics and Beyond*, WIKIBON (Feb. 5, 2014, 3:04 PM), http://wikibon.org/wiki/v/Big_Data:_Hadoop,_Business_Analytics_and_Beyond.

48. *See* Davenport, *supra* note 47, at 66–67.

we change the essence."[49] Thus, big data is about the "moving picture" predictions from unplanned secondary uses of data sets as opposed to earlier eras of planned data processing "snap shots." For example, early pioneers of big data were able to attract viewers to "their websites through better search algorithms, recommendations from friends and colleagues, suggestions for products to buy, and highly targeted ads, all driven by analytics rooted in enormous amounts of data."[50]

We are now entering a third era in which big data use is expanding beyond Silicon Valley innovators to corporate and government institutions. Currently, the volume and variety of data are in ample supply. And it is clear that some of the data we collect today will have unforeseen uses (and value) in the future. These unforeseen secondary uses of data create the incentive for institutions to collect and store data in order to have them for later analysis. Storage, after all, is getting much cheaper, too. Although employees with big data skills have been in relatively short supply,[51] and companies are still learning what to do with big data,[52] this is rapidly changing.

Companies already have access to extensive data sets prepared by a large data broker industry which itself has substantial big data capabilities. The data-driven marketing economy, of which data brokers are a central part, generates revenue in the hundreds of billions of dollars.[53] To obtain their information, data brokers search through government records, purchase histories, social media posts, and hundreds of other available sources. Data brokers compile this information and use it to build comprehensive data profiles about us, all of which they sell in turn to retailers, advertisers, private individuals, nonprofit organizations, law enforcement, and other government agencies.[54]

---

47. MAYER-SCHÖNBERGER & CUKIER, *supra* note 5, at 10.

50. Indraneel Kripabindu Sen Gupta, *Big Data Analysis 3.0—Series 1*, INVISIBLE ANALYSIS (Jan. 12, 2014), http://ianalysis.blogspot.com/2014/01/big-data-analysis-30-series-1.html.

51. *See* Thor Olavsrud, *How to Close the Big Data Skills Gap by Training Your IT Staff*, CIO (Oct. 2, 2013), http://www.cio.com/article/740818/How_to_Close_the_Big_Data_Skills_Gap_by_Training_Your_IT_Staff?page=1&taxonomyId=600010 (discussing the big data skills gap).

52. Matt Asay, *Gartner on Big Data: Everyone's Doing It, No One Knows Why*, READWRITE (Sept. 18, 2013), http://readwrite.com/2013/09/18/gartner-on-big-data-everyones-doing-it-no-one-knows-why#awesm=~orVsnL0seNLQWz.

53. *See* Katy Bachman, *Big Data Added $156 Billion in Revenue to Economy Last Year*, ADWEEK (Oct. 14, 2013, 9:17 AM), http://www.adweek.com/news/technology/big-data-added-156-billion-revenue-economy-last-year-153107 (reporting on a study that estimated "the data-driven market economy added $156 billion in revenue to the U.S. economy" in 2012).

54. U.S. GOV'T ACCOUNTABILITY OFFICE, CONSUMER PRIVACY FRAMEWORK NEEDS TO REFLECT CHANGES IN TECHNOLOGY AND THE MARKETPLACE 2–4 (2013).

On top of these already powerful and highly capable data brokers, innovative and rapidly growing startups are further enhancing data analysis and sharing velocity. Take Palantir, a company that applies antifraud techniques developed at PayPal for antiterrorism.[55] Since its founding in 2004, Palantir has raised $650 million in capital and is purportedly worth $9 billion after its most recent capital raise in 2013.[56] Palantir started as a government contractor for law enforcement and intelligence agencies and is now expanding to pharmaceutical and banking sectors.

The increasing adoption of big data is such that all kinds of human activity, ranging from dating[57] to hiring,[58] voting,[59] policing,[60] and identifying terrorists, have already become heavily influenced by big data techniques. These new insights and predictions are already starting to have an impact on the relationships between citizens, governments, and companies. And it is happening so quickly that most people are not aware of both the scale and the speed of these transformations.

C. *Big Data Awareness*

The Big Data Revolution is fundamentally about awareness. The analysis of relevant big data sets gives us greater awareness of the world that lets us make predictions and solve problems. Take the problem of traffic congestion. One way to map a city's daily traffic flows and congestion might be to let researchers run analytics on cellphone signal logs over a metropolitan area over a long enough period of time to see patterns. In 2012, MIT and UC Berkeley

---

55. *What We Do*, PALNTIR, http://www.palantir.com/what-we-do/ (last vistited May 6, 2014).

56. Reed Albergotti, *Palantir: Big Data, Big Dollars*, WALL ST. J., Dec. 6, 2013, at B5.

57. *See, e.g.*, Jonah Lehrer, *The Web's Cockeyed Cupids*, WALL ST. J. (Mar. 16, 2012, 6:04 PM), http://online.wsj.com/news/articles/SB10001424052702304537904577277830191481536.

58. *See, e.g.*, Don Peck, *They're Watching You at Work*, ATLANTIC (Nov. 20, 2013, 9:07 PM), http://www.theatlantic.com/magazine/archive/2013/12/theyre-watching-you-at-work/354681/.

59. *See, e.g.*, Sasha Issenberg, *How President Obama's Campaign Used Big Data to Rally Individual Voters, Part 1*, MIT TECH. REV. (Dec. 16, 2012), http://www.technologyreview.com/featuredstory/508836/how-obama-used-big-data-to-rally-voters-part-1/ ("The [Obama] campaign didn't just know who you were; it knew exactly how it could turn you into the type of person it wanted you to be.").

60. *See, e.g.*, Jordan Robertson, *How Big Data Could Help Identify the Next Felon—Or Blame the Wrong Guy*, BLOOMBERG (Aug. 15, 2013, 12:01 AM), http://www.bloomberg.com/news/2013-08-14/how-big-data-could-help-identify-the-next-felon-or-blame-the-wrong-guy.html; *see also* Andrew V. Papachristos & Christopher Wildeman, *Network Exposure and Homicide Victimization in an African American Community*, 104 AM. J. PUB. HEALTH 143, 143 (2014) (arguing that awareness of offenders' positions in social networks is "essential to understanding individual victimization within high-risk populations").

researchers did exactly that by analyzing mobile phone traffic logs from cell tower interactions of 680,000 Boston-area commuters.[61] This allowed the researchers to "trace each individual's commute, anonymously, from origin to destination,"[62] and enabled the authors of the study to produce "one of the most detailed maps of urban traffic patterns ever constructed"[63] and uncover "*previously hidden patterns* in urban road usage."[64]

Consider also the problem of terrorism. We live in a time when terrorist attacks are also "previously hidden patterns" until they occur. Big data presents an alluring silver bullet to defend against terrorist attacks by greatly expanding the situational awareness of our security services. Situational awareness has long been a cornerstone of military and emergency response theory.[65] Addressing the lack of awareness of September 11th attackers, Congress passed a series of laws including section 515 of the Homeland Security Act, which requires the National Operations Center to "provide situational awareness and a common operating picture for the entire Federal Government . . . and [to] ensure that critical terrorism and disaster-related information reaches government decision-makers."[66] The law defines the term "situational awareness" as "information gathered from a variety of sources that, when communicated to emergency managers and decision makers, can form the basis for incident management decisionmaking."[67]

Big data takes situational awareness to a new level (at least in theory) by allowing the government to see first, decide first, and act first inside an adversary's decision cycle. This "merely" requires the government to collect everything in advance so that it can search for what it needs when it needs it. After the fact, investigators can identify suspected terrorists if they have access to the big metadata computer's pre-attack data to find signals and inform situational awareness. Thus, in the wake of the Boston Marathon bombing,

---

61. Pu Wang et al., *Understanding Road Usage Patterns in Urban Areas*, 2 NATURE SCI. REP. 1, 1 (2012), *available at* http://www.nature.com/srep/2012/121220/srep01001/pdf/srep01001.pdf.

62. Kevin Hartnett, *Traffic: Which Boston-Area Neighborhoods Are to Blame?*, BOS. GLOBE (Feb. 17, 2013), http://www.bostonglobe.com/ideas/2013/02/17/traffic-which-boston-area-neighborhoods-are-blame/h5qqR3CrHDM3xCNsTqdYxH/story.html.

63. Homeland Security Act of 2002, Pub. L. No. 107-296, § 515, 116 Stat. 2135 (*amended by* Department of Homeland Security Appropriations Act, Pub. L. No 109-295, 120 Stat. 1355, 1409 (2006)) (codified at 6 U.S.C. § 321d(b)(1)-(2) (2012)).

64. Wang, *supra* note 61 (emphasis added).

65. *See, e.g.*, PAUL M. SALMON ET AL., DISTRIBUTED SITUATION AWARENESS: THEORY, MEASUREMENT AND APPLICATION TO TEAMWORK (2009).

66. Homeland Security Act of 2002 § 515, 6 U.S.C. § 321d(b)(1)–(2) (2012).

67. *Id.* § 321d(a).

federal officials accessed Boston cell tower traffic logs much like the researchers discussed earlier, but this time to cross check against surveillance video and eyewitness photography in order to identify the culprits of the Boston Marathon bombing.[68] They also used tools like the one from Topsy labs—recently acquired by Apple[69]—that let officials access the metadata built into every tweet sent in Boston since July 2010 that contained the word "bomb."[70]

More ambitiously, how can security services identify and catch terrorists before they attack? One way would be to let government agencies have the metadata of everything in advance so they can "seed"[71] a database with identifiers, such as phone numbers. Such a tactic would have the potential to uncover hidden patterns that could help analysts combine with other sources of intelligence to determine if an attack was about to happen. Big data analytics could also allow the identification of groups of suspected terrorists once the identity of their phone numbers became known. Internationally, this could take the form of allowing the NSA to collect global data on all cellular traffic it could possibly access and correlate the data of who is calling whom, how often, and when certain numbers are at certain locations and times when certain indicators are present.[72] Domestically, we could also allow the NSA to collect metadata from domestic carriers and store them in one historical depository that it could keep for a fixed period (say, five years) and that it could retrospectively query to "discern connections between terrorist organizations and previously unknown terrorist operatives located in the United States."[73] In fact, something like this is happening as this Article is going to press with President Obama's proposal for reform legislation that would instead keep bulk phone call data with telephone companies.[74]

Big data will increasingly inform everyday policing and cyber security efforts. Law enforcement of all kinds, state and local, are making use of big data practices to pinpoint potential crime hot

68. *See* Frank Konkel, *Boston Probe's Big Data Use Hints at the Future*, FCW (Apr. 26, 2013), http://fcw.com/articles/2013/04/26/big-data-boston-bomb-probe.aspx.

69. Daisuke Wakabayashi & Douglas Macmillan, *Apple Taps into Twitter, Buying Social Analytics Firm Topsy*, WALL ST. J. (Dec. 2, 2013, 9:30 PM), http://online.wsj.com/news/articles/SB10001424052702304854804579234450633315742.

70. *See* Konkel, *supra* note 68.

71. *See* Klayman v. Obama, 957 F. Supp. 2d 1, 16 (D.D.C. 2013).

72. Barton Gellman & Ashkan Soltani, *NSA Maps Targets by Their Phones*, WASH. POST, Dec. 5, 2013, at A1.

73. *Klayman*, 957 F. Supp. 2d at 15.

74. *See* Charlie Savage, *Obama to Call for End to N.S.A.'s Bulk Data Collection*, N.Y. TIMES, Mar. 25, 2014, at A1.

spots or predict houses that could be burglarized.[75] Some experimental departments are even developing algorithms to predict future felons.[76] Cyber attacks of all kinds are on the rise. One way to defend against these attacks is to use big data to become aware of cyber attacks and to find vulnerabilities to defend against an attack. With the threats posed by cyber attacks, both government agencies like the NSA and corporations like Microsoft[77] will need to be prepared to act in the big metadata computer in a much more pervasive, persistent, and invasive way because they need to protect the big metadata computer itself.

On the one hand, it should be no surprise that companies and governments are aggressively mobilizing big data to improve products and defend against terrorist and cyber attacks. On the other hand, it should be no surprise that the public is starting to ask questions about privacy as it learns about the potential privacy invasions that big data awareness allows. Yet many of the problems that concern us about big data extend beyond narrow notions of privacy. We worry about our confidential information being disclosed to unknown third parties. Moreover, we lack the transparency needed to gauge the effect of big data predictions and inferences upon us because the operations of big data themselves are shrouded in legal and commercial secrecy. As we start to learn about surprising uses of this shared information, we wonder how it may change who we are, for the better or for the worse. As the facts surrounding actual uses of big data continue to emerge, we are in a critical window before mass big data adoption where we can develop principles to capture the promise of big data without losing important societal values.

II. BIG DATA ETHICS

We are living in a time when new kinds of information collection and analysis promise great things, especially by increasing our awareness about society. And when it comes to awareness about the people who make up our society, the Big Data Revolution is being recorded by what we might think of as a "big metadata computer," comprised of data about people and metadata about that data. We have some privacy rules to govern existing flows of personal information, but we lack rules to govern new flows, new uses, and new decisions derived from that data. What we need

---

75. *See, e.g.*, Kevin Fogarty, *Big Data Plus Police Work: Good Partners?*, INFO. WK. (July 24, 2012, 3:36 PM), http://www.informationweek.com/software/information-management/big-data-plus-police-work-good-partners/d/d-id/1105482.

76. Robertson, *supra* note 60.

77. *See* Matthew J. Schwartz, *Microsoft, FBI Trumpet Citadel Botnet Takedowns*, INFO. WK. (June 6, 2013, 10:26 AM), http://www.informationweek.com/attacks/microsoft-fbi-trumpet-citadel-botnet-takedowns/d/d-id/1110261.

are new rules to regulate the societal costs of our new tools without sacrificing their undeniable benefits.

But what values should guide us in forming these new rules? In this Part, we argue that a set of four normative values (privacy, confidentiality, transparency, and identity) suggests the beginnings of "Big Data Ethics" to govern data flows in our information society and inform the establishment of legal and ethical big data norms.

A. Privacy

We typically think about problems of personal information under the rubric of "privacy." But the Big Data Revolution need not signal the "death of privacy." On the contrary, when we think of "privacy" as more than keeping secrets and recognize it instead as the rules we have to govern information flows, big data's real privacy problem comes into focus. We need rules to regulate the flows of data, which means that the collection of personal data should be the beginning of our privacy conversation and not its end.

1. Privacy as Information Rules

We are lured to think that the Big Data Revolution will eliminate privacy when many of its leading proponents declare that "Privacy is dead" or "Privacy is dying." In January 1999, Sun Microsystems CEO Scott McNealy famously declared, "You have zero privacy anyway. . . . Get over it."[78] McNealy's outburst made headlines at the time, and it has outlived Sun's own existence as an independent company. More recently, Vint Cerf, a leading figure in the creation of the Internet and Google's "Chief Internet Evangelist," suggested that privacy might be a historical anomaly.[79] Facebook founder Mark Zuckerberg was more blunt, declaring that "the age of privacy is over."[80] Such techno-centric worldviews carry an implied undertone of technology infallibility. We must yield our expectations of privacy, they suggest, to make way for the inevitable, and get out of the way of technological innovation.

Yet Edward Snowden and Glenn Greenwald's revelations about the scale of surveillance by the National Security Agency have prompted a global debate about surveillance and privacy that continues months later. Why is this happening if privacy is dead? We would like to suggest, to the contrary, that *privacy is not dead.*

---

78. Polly Sprenger, *Sun on Privacy: "Get Over It,"* WIRED (Jan. 26, 1999), http://archive.wired.com/politics/law/news/1999/01/17538.

79. Gregory Ferenstein, *Google's Cerf Says "Privacy May Be An Anomaly." Historically, He's Right.*, TECHCRUNCH (Nov. 20, 2013), http://techcrunch.com/2013/11/20/googles-cerf-says-privacy-may-be-an-anomaly-historically-hes-right/.

80. Marshall Kirkpatrick, *Facebook's Zuckerberg Says the Age of Privacy Is Over*, READWRITE (Jan. 9, 2010), http://readwrite.com/2010/01/09/facebooks_zuckerberg_says_the_age_of_privacy_is_ov#awesm=~oo2UUoqssyO3eq.

Privacy is very much alive, though it, like other social norms, is in a state of flux.

It all depends on what we mean by "privacy." If we think about privacy as the amount of information we can keep secret or unknown, then that kind of privacy is certainly shrinking. We are living through an information revolution, and the collection, use, and analysis of personal data is inevitable. But if we think about privacy as the question of what rules should govern the use of personal information, then privacy has never been more alive. In fact, it is one of the most important and most vital issues we face as a society today.

Our definitions of privacy matter. A simplistic definition of privacy that is often used in public debates is something like "the information about me that no one knows." But lawyers have understood privacy in more sophisticated ways for decades. At a minimum, lawyers use the word "privacy" and the legal rules that govern it to mean four discrete things: (1) invasions into protected spaces, relationships, or decisions; (2) collection of information; (3) use of information; and (4) disclosure of information.[81] In the leading conceptual work on privacy, legal scholar Daniel Solove has taken these four categories and expanded them to *sixteen* categories, including surveillance, interrogation, aggregation, and disclosure.[82]

Though we will need new privacy rules for the many uses of information, as the Information Revolution develops, we have many such rules already. Some of these rules are ones that we typically think of as "privacy rules." For example, tort law governs invasions of privacy including peeping (or listening) Toms,[83] the unauthorized use of photographs for commerce,[84] and the disclosure of sexual images without consent.[85] The Fourth Amendment requires that the government obtain a warrant before it intrudes on a "reasonable expectation of privacy," and a complex web of federal and state laws regulating eavesdropping and wiretapping by both government and private actors backs up the Fourth Amendment.[86] In addition to the Privacy Act and the Fair Credit Reporting Act, federal laws regulate the collection and use of financial information, medical and genetic

---

81. *Cf.* Neil M. Richards, *Reconciling Data Privacy and the First Amendment*, 52 UCLA L. REV. 1149, 1181–82 (2005) (categorizing the regulation of information into four similar categories).

82. DANIEL J. SOLOVE, UNDERSTANDING PRIVACY 10–11 (2008).

83. *See generally* Hamberger v. Eastman, 206 A.2d 239, 241–42 (N.H. 1964).

84. *See* RESTATEMENT (SECOND) OF TORTS § 652C (1977).

85. *See generally* Michaels v. Internet Entm't Grp., 5 F. Supp. 2d 823, 840–42 (C.D. Cal. 1998).

86. *See, e.g.*, Electronic Communications Privacy Act of 1986, 18 U.S.C. §§ 2510–2522 (2012); CAL. PENAL CODE § 632(a) (Deering 2008); Katz v. United States, 389 U.S. 347, 357–58 (1967).

information, and video privacy, among others.[87] States, led by California, have also added privacy protections, such as California's constitutional right of privacy (applicable to private actors), reading privacy laws, data breach notification statutes, and the recent spate of laws prohibiting employers from asking for the social media account passwords of their employees.[88] Even the First Amendment, long thought of as the enemy of privacy, is a kind of information rule that mandates the circumstances in which other laws cannot restrict certain free flows of information, such as the publication of true and newsworthy facts by journalists, or truthful and nonmisleading advertisements for lawful products.[89]

The important point we want make here is this: however we define privacy, it will have to do with information. Privacy should not be thought of merely as how much is secret, but rather about what rules are in place (legal, social, or otherwise) to govern the use of information as well as its disclosure. The law has actually thought of privacy in this way for a very long time in a number of ways, including, for example, in the protection of confidences.[90] And when we think of information rules as privacy rules, we can see that even though digital technologies and government and corporate practices are putting many existing notions of privacy under threat, privacy in general is not dying. This is because privacy is more than just secrecy. Privacy is a shorthand we have come to use to identify information rules. As Helen Nissenbaum has put it, when we talk about privacy, we mean the rules that govern how information flows and not merely restrictions on acquiring personal information or data.[91]

If we were designing things from scratch, we would almost certainly want to use a word other than "privacy"; "information rules" springs to mind, as does the more accurate but less exciting European concept of "data protection." But in the English-speaking world at least, "privacy" is so deeply rooted as the word we use to

87. *See generally* Privacy Act of 1974, 5 U.S.C. § 552a (2012); Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681 (2012); Gramm-Leach-Bliley Act, 15 U.S.C. §§ 6801–6809 (2012); Video Privacy Protection Act of 1988, 18 U.S.C. §§ 2701–2712 (2012); Health Insurance Portability and Accountability Act (HIPAA) of 1996, 42 U.S.C. §§ 201–300ii (2012).

88. *E.g.*, CAL. CONST. art. I, § 1; CAL. CIV. CODE § 1798.82 (West 2014) (requiring notification of certain data breaches); Reader Privacy Act, CAL. CIV. CODE § 1798.90 (West 2012); CAL. LABOR CODE § 980 (West 2014) (prohibiting certain employer actions with regard to social media).

89. *See generally* Neil M. Richards, *Why Data Privacy Law Is (Mostly) Constitutional* (Oct. 2, 2013) (unpublished manuscript), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2335196.

90. *See* Neil M. Richards & Daniel J. Solove, *Privacy's Other Path: Recovering the Law of Confidentiality*, 96 GEO. L.J. 123, 133–38 (2007) (discussing how American law protected personal information from disclosure through confidentiality rules).

91. HELEN NISSENBAUM, PRIVACY IN CONTEXT 1–2 (2010).

refer to the collection, use, and disclosure of information that we are probably stuck with it, for better and for worse. When we expand our idea of "privacy" beyond embarrassing secrets to include the regulation of information, it flows more generally, and we see that privacy—and privacy law—is imperative in today's information economy.

The "death of privacy" really refers to two somewhat related phenomena. First, there is the phenomenon of large amounts of personal information being collected by the technologies that we lump together metaphorically as the "big metadata computer" in Part I. But since privacy means more than protection from collection, the fact that we have big data *increases* the need for and importance of privacy rules, rather than decreasing it. It does seem to be true that social expectations about shared information are changing. But our social understandings about lots of things (including privacy) are always in flux. Moreover, the legal and social rules that govern how information about us is obtained and used (broadly defined) are always necessary, and the Information Revolution is increasing the importance of these information rules rather than decreasing it.

Second, and just as important, if there is a sense of a crisis in personal information, what has broken is not our concern about information rules or the need for them but our practical ability as individuals to manage the trade in and uses of information about us. Existing privacy law focuses on a set of principles known as the "Fair Information Principles" to govern the collection, use, and disclosure of personal data.[92] The objective is to provide individuals control over their personal data so that they can weigh the benefits and costs at the time of collection, use, or disclosure. And the most important principles in practice as the law has evolved are notice (the idea that data processors should disclose what they are doing with personal data) and choice (the idea that people should be able to opt-out of uses of their data that they dislike). The "notice and choice" regime is the basic framework on which our current system of privacy policies, privacy settings, and privacy dashboards operates.

Professor Daniel Solove describes this approach to privacy regulation as "privacy self-management."[93] While privacy self-management promises nuanced privacy protection, in practice most companies provide constructive notice at best, and individuals make take-it-or-leave-it decisions to provide consent.[94] Few individuals, if

92. *See* DANIEL J. SOLOVE & PAUL M. SCHWARTZ, INFORMATION PRIVACY LAW 698–700 (4th ed. 2011).

93. Solove, *supra* note 1, at 1880.

94. Paul M. Schwartz, *Beyond Lessig's Code for Internet Privacy: Cyberspace Filters, Privacy Control, and Fair Information Practices*, 2000 WIS. L. REV. 744, 768–69.

any, have the time, skill, or dedication to go through the complex web of terms and conditions of any given consent, let alone revisit consent once given.[95] Indeed, empirical evidence suggests that privacy self-management of the sort the model expects (reading privacy policies and making granular choices) would take users hundreds of hours per year to actually accomplish.[96] Solove sums this up as a consent dilemma where many recent attempts at reform really just call for more privacy self-management, and alternative paternalistic approaches just limit individual consent.[97] The problem is not that privacy is dead but rather that the system of managing the flows of personal information needs to be rethought in the face of the new uses and sources that our Information Revolution has generated. The problem is thus not the death of privacy but rather the need for additional principles to govern information flows.

2. Shared Private Information Can Remain Confidential

Of course, while the notion of privacy as secrecy is common in public debate, our law has a more nuanced understanding. In the previous section, we argued that we should understand privacy (or at least privacy law) as the set of rules we use to govern the flow of personal information that makes up much of our information economy. We argued that we should reject narrow understandings of privacy, like the understanding that privacy is just about keeping secrets from the world. Such notions of privacy are binary; information in this view is either on or off, public or private, known to us alone or broadcast to the world. But such narrow understandings of privacy are (to be blunt) nonsense. Information is rarely known to all or known to none. Instead, virtually all information exists in intermediate states between completely public and completely private. Much of the information in intermediate states that we share is private data that we share in trust, expecting them to remain confidential. Confidentiality is a kind of privacy that is based on trust and reliance on promises in the context of relationships.[98] With the power of big data to make secondary uses of the private information we share in confidence, restoration of trust in the institutions we share with rests not only with privacy but in the recognition that *shared private information can remain "confidential."* In other words, private digital information that we

95. *See* David Pogue, *Term of Confusion*, SCI. AM., Mar. 2013, at 35 (noting that terms and conditions are often overly complex and difficult for normal people to understand).

96. Aleecia M. McDonald & Lorrie Faith Cranor, *The Cost of Reading Privacy Policies*, 4 I/S: J.L. & POL'Y FOR INFO. SOC'Y 543, 564–65 (2008).

97. Solove, *supra* note 1, at 1881–82.

98. *See* Richards & Solove, *supra* note 90, at 125.

share with third parties we trust can still be regulated by privacy law.

Binary notions of privacy are particularly dangerous in our digital era, where information is necessarily shared by design in order to be more useful. We welcome GPS, cell tower, and even Wi-Fi location tracking of our cell phones so that we can make calls more easily and use location services in applications to "check-in," navigate, or find our friends. We willingly share information to feed big data algorithms so dating sites can find us compatible mates, career sites can help us more quickly find jobs, online bookstores can recommend books for us to read, and social networking sites can connect us with new friends. Yet as discussed in Part I, the information we generate lives on and the emergence of big data allows for increased insights that can create digital dossiers about us that we know little, if anything, about.[99] Before big data, individuals could roughly gauge the expected uses of their personal data and weigh the benefits and the costs at the time they provided their consent. Even if they guessed wrong, they would have some comfort that the receiving party would not be able to make additional use of their personal data.[100] The growing adoption of big data and its ability to make extensive, often unexpected, secondary uses of personal data changes this calculus. As Kord Davis observed in his book *Ethics of Big Data*, "the potential for harm due to unintended consequences, can quickly outweigh the value the big-data innovation is intended to provide."[101] Not only is privacy self-management broken, but these new technological advances will compound the harm that comes from its failure.

These unintended consequences may not only involve individual privacy, they may also cause substantial harm to institutions. One particularly important injury is the loss of trust. Since the Edward Snowden revelations about NSA spying were first published in June 2013, the U.S. government has been managing through a kind of trust outage with untold cost to taxpayers and its mission.[102] General Alexander acknowledged the loss of trust impacting the NSA's cyber mission. "'Cyber is where we need allies and partners around the world,' Alexander said. 'In order to get there, we need to

99. *See* DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 2 (2004).

100. *See* Woodrow Hartzog & Evan Selinger, *Big Data in Small Hands*, 66 STAN. L. REV. ONLINE 81, 84 (2013), http://www.stanfordlawreview.org/online/privacy-and-big-data/big-data-small-hands.

101. *See* KORD DAVIS & DOUG PATTERSON, ETHICS OF BIG DATA 5 (2012).

102. Steve Rosenbush, *Obama Addresses Economic Damage Caused by Snowden NSA Leaks*, CIO J. (Jan. 17, 2014, 5:44 PM), http://blogs.wsj.com/cio/2014/01/17/obama-addresses-economic-damage-caused-by-snowden-nsa-leaks/.

change the rhetoric on media leaks, and fix the trust factor.'"[103] The NSA's trust outage is not only hurting itself but also entire sectors of the United States' information technology industry, as foreign countries both react to protect their citizens' privacy and use the trust outage as a means to advance local competitors.[104]

Technology companies have also called for the restoration of trust. On October 9, 2013, Apple, Google, Microsoft, Facebook, Yahoo, LinkedIn, Twitter, and AOL published an open letter to President Barack Obama and Congress calling for surveillance reform.[105] Brad Smith, Microsoft's general counsel, wrote, "People won't use technology they don't trust. Governments have put this trust at risk, and governments need to help restore it."[106] Marissa Mayer, CEO of Yahoo, argued that "[r]ecent revelations about government surveillance activities have shaken the trust of our users, and it is time for the United States government to act to restore the confidence of citizens around the world."[107] These technology providers call for reform because they fear that customers will lose trust in their services if their customers' shared private data is no longer confidential.

Confidentiality law arose centuries ago to keep certain kinds of shared information private.[108] Multiple areas of the law provide confidentiality protections for preventing the disclosure of information in intermediate states, whether through professional duties of confidentiality, implied or expressed contracts for confidentiality, evidentiary privileges, or statutory rules.[109] We have long had confidentiality rules like the duties lawyers owe to their clients and doctors owe to their patients to incent individuals to feel safe in sharing their confidences to advance important societal values of providing effective legal representation and medical care.[110] We also have statutory rules that explicitly create

---

103. *See* Grant Gross, *NSA's Alexander to Telecom Industry: Trust Me*, PCWORLD (Oct. 9, 2013, 10:57 AM), http://www.pcworld.com/article/2053540/nsas-alexander-asks-telecom-industry-to-trust-him.html.

104. *See* James Staten, *The Cost of Prism Will Be Larger than ITIF Projects*, FORRESTER BLOGS (Aug. 14, 2013), http://blogs.forrester.com/james_staten/13-08-14-the_cost_of_prism_will_be_larger_than_itif_projects.

105. *Global Government Surveillance Reform*, REF. GOV'T SURVEILLANCE, http://reformgovernmentsurveillance.com (last visited Apr. 8, 2014) (displaying an open letter from AOL, Apple, Facebook, Google, LinkedIn, Microsoft, Twitter & Yahoo to Washington).

106. *Id.*

107. *Id.*

108. *See* Richards & Solove, *supra* note 90 (tracing the history of confidentiality law).

109. *See* Woodrow Hartzog, *Chain-Link Confidentiality*, 46 GA. L. REV. 657, 668–75 (2012) (providing a comprehensive review of confidentiality law).

110. *See, e.g.*, Upjohn Co. v. United States, 449 U.S. 383, 389 (1981).

confidential relationships regarding health,[111] financial,[112] and video records[113] information. We also protect obligations of confidentiality that arise through voluntary promises or confidentiality agreements like preventing employees from revealing business secrets.[114] Confidentiality law reveals how we have long recognized shared information can still be kept private using effective legal tools. Expanding confidentiality law approaches would seem to be one way to help keep shared information private.

Another force helping to keep shared private information confidential is the Federal Trade Commission ("FTC"). Since the late 1990s, the FTC has maintained that breaking promises in a privacy notice constitutes an "unfair or deceptive act" under the Federal Trade Commission Act.[115] The FTC can bring civil actions and seek injunctive remedies when it finds such promises broken.[116] Solove and Hartzog explain how the FTC's privacy jurisprudence has become the functional equivalent of a body of common law for privacy law.[117] They go on to observe that the FTC is now starting to move "beyond the four corners of privacy policies" and shift its focus from enforcing broken promises of privacy to broken expectations of consumer privacy.[118] This subtle but powerful shift puts the FTC in a position to increasingly look at the totality of circumstances surrounding privacy policies, including when consumers assume their shared information is being kept private. This expanded view could put the FTC in a position to "demand that companies engage in practices that will correct mistaken consumer assumptions or at the very least not exploit such assumptions," like when consumers assume their shared private information is being kept confidential.[119]

Courts are also starting to grapple with the privacy expectations surrounding a new kind of shared private information to keep confidential—metadata. While much of the actual data

---

111. *See, e.g.*, Health Insurance Portability and Accountability Act (HIPAA) of 1996, 42 U.S.C. §§ 1320d–1320d-8 (2012) (regulating the disclosure of information related to individuals' health care).

112. *See, e.g.*, Fair Credit Reporting Act (FCRA), 15 U.S.C. § 1681 (2012); Gramm-Leach-Bliley Act, 15 U.S.C. §§ 6801–6809.

113. Video Privacy Protection Act of 1988, 18 U.S.C. § 2710(b)(2)(B) (2012).

114. *See, e.g.*, Raven Indus. v. Lee, 783 N.W.2d 844, 847–851 (S.D. 2010) (enforcing an employee nondisclosure agreement).

115. 15 U.S.C. § 45; *see* Marcia Hoffman, *Federal Trade Commission Enforcement of Privacy*, *in* PROSKAUER ON PRIVACY: A GUIDE TO PRIVACY AND DATA SECURITY LAW IN THE INFORMATION AGE 4-1 (Kristen J. Mathews ed., 2013).

116. *See* Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. (forthcoming 2014), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2312913.

117. *Id.* at 1.

118. *Id.* at 57.

119. *Id.* at 56.

collected and stored has some kind of protection surrounding it, the associated metadata often does not. For example, the Electronic Communications Privacy Act ("ECPA") prevents Internet service providers from selling the content of its customers' e-mails and text messages without written consent but provides more limited protection for noncontent metadata.[120] This is out of touch with today's world because metadata are being created more easily than ever and can be aggregated with other information to reveal as much or more about individuals as personally identifying information or actual data.[121] Moreover, metadata are often easier to access and share and can enable de-identification, allowing for even more privacy and identity intrusion.[122]

Some courts and several state legislatures[123] are starting to recognize the privacy implications of metadata collection. In *Klayman v. Obama*, Judge Richard Leon granted and then stayed, for national security reasons, a preliminary injunction to stop the government's bulk collection and querying of the plaintiff's phone record metadata on Fourth Amendment grounds.[124] The government argued that based on the Supreme Court's ruling in *Smith v. Maryland*[125] in 1979, "no one has an expectation of privacy, let alone a reasonable one, in the telephony metadata that telecom companies hold as business records."[126] Judge Leon distinguished *Smith* by framing the question in *Klayman* as: "When do present-day circumstances—the evolutions in the Government's surveillance capabilities, citizens' phone habits, and the relationship between the NSA and telecom companies—become so thoroughly unlike those considered by the Supreme Court thirty-four years ago that a precedent like *Smith* simply does not apply?"[127]

Judge Leon relied in part on the Supreme Court's recent decision in *United States v. Jones,* where the majority ruled based on a trespass rationale that the government's installation of a GPS device on a target's vehicle, and its use of that device to monitor the vehicle's movement for longer than the original warrant period, constituted a search.[128] Justice Sotomayor concurred with the trespass rationale of the majority in *Jones* but went on to observe

120. *See* 18 U.S.C. § 2702 (2012).

121. *See* Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 506–07 (2006).

122. *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1705 (2010); Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a New Concept of Personally Identifiable Information*, 86 N.Y.U. L. REV. 1814, 1879–83 (2011).

123. *See, e.g.*, MONT. CODE ANN. § 46-5-110 (2013).

124. *See* Klayman v. Obama, 957 F. Supp. 2d 1, 43 (D.D.C. 2013).

125. 442 U.S. 735, 745–46 (1979).

126. *Klayman*, 957 F. Supp. 2d at 31.

127. *Id.*

128. *See* United States v. Jones, 132 S. Ct. 945, 949 (2012).

that GPS metadata "generates a precise, comprehensive record of a person's public movements that reflects a wealth of detail about her familial, political, professional, religious, and sexual associations."[129] Sotomayor worried presciently that "[t]he Government can store such records and efficiently mine them for information years into the future."[130] More broadly, Sotomayor questioned the underlying premise "that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties."[131] Justice Sotomayor observed that in the digital age, "people reveal a great deal about themselves to third parties in the course of carrying out mundane tasks."[132]

But the matter of metadata in the courts is far from settled. Just two weeks after Judge Leon's ruling in *Klayman*, U.S. District Court Judge William H. Pauley III did not distinguish *Smith* and ruled that the government's bulk metadata program did not violate the Fourth Amendment.[133] Addressing location metadata in July 2013, the U.S. Court of Appeals for the Fifth Circuit ruled that information revealed by cell phone tower records is not something in which individuals have a "reasonable expectation of privacy."[134] The court reasoned that "[a] cell service subscriber, like a telephone user, understands that his cell phone must send a signal to a nearby cell tower in order to wirelessly connect his call."[135] Since no physical intrusion occurred as in *Jones*, the police could monitor warrant-free according to the Stored Communications Act.[136]

Judges excluding metadata or information shared in trust from "reasonable expectations of privacy" rulings repeat the mistakes of technology leaders who spread "privacy is dead" myths. Limited expectations of privacy rulings and the administration's reliance upon them perpetuate limited expectations of privacy. This causes confusion and delay in responsibly aligning our laws to realize the full benefits of the Big Data Revolution we are privileged to be living in. For example, continued reliance on the thirty-four-year-old *Smith* ruling, based on a collection of information on one phone line on one person for a limited period of time, somehow became the justification for all three branches to justify the collection of nearly

129. *Id.* at 955 (Sotomayor, J., concurring).
130. *Id.* at 955–56.
131. *Id.* at 957.
132. *Id.*
133. *See* ACLU v. Clapper, 959 F. Supp. 2d 724 (S.D.N.Y. 2013).
134. *In re* Application of the U.S. for Historical Cell Site Data, 724 F.3d 600, 608, 615 (5th Cir. 2013).
135. *Id.* at 613.
136. *See* Neil M. Richards, *They Know Where You Are (but They Shouldn't)*, BOS. REV. (Aug. 6, 2013), https://www.bostonreview.net/blog/they-know-where-you-are-they-shouldn't.

every American's phone metadata record for seven years.[137] Even Stephen Sachs, the distinguished Maryland Attorney General who argued and won *Smith*, believes that "the circumstances are radically different today.... To extend it to what we now know as massive surveillance, in my personal view, is a bridge too far."[138]

Fundamentally, the debate over *Smith v. Maryland*'s "third party doctrine" is one about definitions of privacy.[139] The government asserts that once information is shared it can no longer be protected. Such a bald assertion is inconsistent with both the needs of the information age and with common sense. Longstanding legal principles of confidentiality show the way forward, that when appropriate, we can protect private information that exists in intermediate states. Paradoxically, confidentiality provides the trust necessary to ensure that better sharing takes place under terms that are clear, allowing the benefits of sharing and the protection of privacy at the same time.

3. Transparency

Transparency, like confidentiality, also fosters trust by being able to hold others accountable. Transparency of government information plays a crucial role in ensuring constitutional checks and balances among the branches of government, a free press, and individual citizens.[140] Transparency of financial reporting fuels investors' willingness to part with their money and buy stocks. To hold the government accountable, Congress enacted the Freedom of Information Act in 1966 to enable transparent access of information to individuals and companies without the need for a reason.[141] Recognizing the need for transparency, the Obama administration issued several memoranda on transparency and open government as soon as it took office.[142] The European Union Data Protection

137. *See* Klayman v. Obama, 957 F. Supp. 2d 1, 32 (D.D.C. 2013); David Kravets, *How a Purse Snatching Led to the Legal Justification for NSA Domestic Spying*, WIRED (Oct. 2, 2013, 6:30 AM), http://www.wired.com/2013/10/nsa-smith-purse-snatching/.

138. Kravets, *supra* note 137.

139. Eric Smith Dennis, Note, *A Mosaic Shield:* Maynard, *the Fourth Amendment, and Privacy Rights in the Digital Age*, 33 CARDOZO L. REV. 737, 749 (2011).

140. *See* Sidney A. Shapiro & Rena I. Steinzor, *The People's Agent: Executive Branch Secrecy and Accountability in an Age of Terrorism*, 69 LAW & CONTEMP. PROBS. 99, 128 (2006).

141. *See* Freedom of Information Act, Pub. L. No. 89-487, 80 Stat. 250 (1996) (codified as amended at 5 U.S.C. § 552 (2012)).

142. *See* Memorandum on the Freedom of Information Act, 2009 DAILY COMP. PRES. DOC. 9 (Jan. 26, 2009); Memorandum on Transparency and Open Government, 2009 DAILY COMP. PRES. DOC. 10 (Jan. 21, 2009); OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, MEMORANDUM ON OPEN GOVERNMENT DIRECTIVE (Dec. 8, 2009), *available at* http://www.whitehouse.gov/omb/assets/memoranda_2010/m10-06.pdf.

Directive already provides transparency protections.[143] And Ira Rubenstein, Doc Searls, and others describe a future where additional transparency protections will allow data portability to support new business models to enable consumers control over their personal data.[144] Alex Pentland, in his book *Social Physics*, proposes a "New Deal on Data" that would provide enhanced tools for privacy and transparency to allow the use of personal data "to both build a better society and to protect the rights of the average citizen."[145]

Transparency inherently includes a tension between openness and secrecy. This tension can cause paradoxes. Transparency of sensitive corporate or government secrets could harm important interests, such as trade secrets or national security. Too little transparency can lead to unexpected outcomes and a lack of trust. Transparency also carries the risk that inadvertent disclosures will cause unexpected outcomes that harm privacy and breach confidentiality.[146]

In our last paper, we described a "Transparency Paradox" of big data where all manner of data is collected on individuals by institutions while these same institutions are cloaked in legal and commercial secrecy.[147] In order to carry out their mission or provide their services, government agencies like the NSA and companies like Facebook use suites of robust legal tools to preserve their own privacy. Yet, at the same time, these institutions demand and shape transparent collection from us, especially where they have institutional incentives to protect government interests or make money. In an added twist, companies like Google, Apple, and Microsoft make demands for governmental transparency[148] to enable them to issue transparency reports while these same

143. Directive 95/46/EC, of the European Parliament and of the Council, 1995 O.J. (L 281) 31, 38 (EC).

144. *See, e.g.*, Ira S. Rubinstein, *Big Data: The End of Privacy or a New Beginning?*, 3 INT'L DATA PRIVACY L. 74, 81 (2013); DOC SEARLS, THE INTENTION ECONOMY: WHEN CUSTOMERS TAKE CHARGE (2012); *see also* Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 242 (2013).

145. ALEX PENTLAND, SOCIAL PHYSICS: HOW GOOD IDEAS SPREAD THE LESSONS FROM A NEW SCIENCE 178 (2014).

146. *See* Shawn Musgrave, *Boston Police Halt License Scanning Program*, BOS. GLOBE (Dec. 14, 2013), http://www.bostonglobe.com/metro/2013/12/14/boston-police-suspend-use-high-tech-licence-plate-readers-amid-privacy-concerns/B2hy9UIzC7KzebnGyQ0JNM/story.html?s_campaign=sm_tw.

147. *See* Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41 (2013), http://www.stanfordlawreview.org/online/privacy-and-big-data/three-paradoxes-big-data.

148. *See* Andrew Couts, *Google, Microsoft, Apple, and More Launch 'Reform Government Surveillance' Campaign*, DIGITAL TRENDS (Dec. 9, 2013), http://www.digitaltrends.com/web/tech-giants-launch-reform-government-surveillance-campaign/.

companies implement sophisticated encryption called "Perfect Forward Secrecy" to make their data less transparent to government snooping.[149] Google, at least, deserves some credit for their recent efforts to advance transparency with the Google Dashboard, which lets individual users know what data Google has about them.[150]

Transparency has heightened importance with the arrival of big data.[151] The power of big data comes in large part from secondary uses of data sets to produce new predictions and inferences. As discussed in Part I, institutions like data brokers, often without our knowledge or consent, are collecting massive amounts of data about us they can use and share in secondary ways that we do not want or expect. Because of this, data brokers have recently come under attack for not meeting many of the "Fair Information Practice" principles ("FIPs"), especially those relating to transparency. In February 2012, the FTC issued a privacy report calling upon Congress to give consumers more control over their information held by data brokers.[152] In December 2012, the FTC launched a privacy probe to study the data broker industry's collection and use of consumer data.[153] In a recent report on the data broker industry, Senator Rockefeller stressed that lack of data broker transparency regarding data sources and use only exacerbates an "aura of secrecy surrounding the industry."[154]

Our point here is not to pick on the data broker industry, but to draw attention to the complexity of privacy in an age in which it is purportedly dead. Rather than having no privacy for individuals and maximal privacy for institutions, we think a better balance is necessary, in which individuals need more privacy and institutions need less. After all, Louis Brandeis himself famously explained that

149. *See* Nicole Perlroth & Vindu Goel, *Internet Firms Step Up Efforts to Stop Spying*, N.Y. TIMES, Dec. 5, 2013, at A1.

150. PENTLAND, *supra* note 145, at 183.

151. *See* Audrey Watters, *What Does Privacy Mean in an Age of Big Data?*, O'REILLY (Nov. 2, 2011), http://strata.oreilly.com/2011/11/privacy-big-data-transparency.html (documenting an interview with author Terence Craig on the importance of transparency in the age of big data).

152. *See* Rainey Reitman, *FTC Final Privacy Report Draws a Map to Meaningful Privacy Protection in the Online World*, ELECTRONIC FRONTIER FOUND. (Mar. 26, 2012), https://www.eff.org/deeplinks/2012/03/ftc-final-privacy-report-draws-map-meaningful-privacy-protection-online-world.

153. *See* Katy Bachman, *FTC Launches Probe of Data Broker Privacy Practices*, ADWEEK (Dec. 18, 2012, 12:30 PM), http://www.adweek.com/news/technology/ftc-launches-probe-data-broker-privacy-practices-146041.

154. *See* Adam Tanner, *Senate Report Blasts Data Brokers for Continued Secrecy*, FORBES (Dec. 19, 2013, 10:00 AM), http://www.forbes.com/sites/adamtanner/2013/12/19/senate-report-blasts-data-brokers-for-continued-secrecy/.

"sunlight . . . is the best of disinfectants."[155] If a big–data governed society is to have any rules, those who collect, share, and use data must be made more transparent and thus more accountable. If we know that companies have the ability to issue transparency reports on government requests for information, we can better trust in the government making the request. Going further, however, if we know these same companies have transparency policies on their own collection, sharing, and usage of data about us, we will have greater confidence in them as well.

B. Identity

Big data requires us also to think more deeply about identity. Identity, like privacy, is hard to define but equally vital to protect. Whereas privacy harkens from the right to be let alone, identity hails from the fundamental right to define who we are. Protecting privacy, especially intellectual privacies, helps protect identity by giving individuals room to make up their own minds.[156] Yet privacy protections are not enough in our new age of the big metadata computer because big data analytics can compromise identity by allowing institutional surveillance to moderate and even determine who we are before we make up our own minds. Therefore, we are concerned that big data can compromise identity and believe that, in addition to privacy and confidentiality protections, we must begin to think about the kinds of big data predictions and inferences that we will allow and the ones that we should not.

Identity can mean many things. It can refer to the association of a specific name to a specific person. Indeed, entire industries of identity management and identity protection now exist to protect this kind of identity. Identity can also mean whether something is the same as something or someone else, as it is treated in evidence law.[157] Philosophers have also long debated and tried to define identity in this fashion. In this debate, the identity of a thing, including a person, is comprised of those properties or qualities which make it that thing. The problem with the philosophical definition of identity is that if you change the properties or qualities of the thing, you no longer have the same thing.[158]

---

155. *See* Neil M. Richards, *The Puzzle of Brandeis, Privacy, and Speech*, 63 VAND. L. REV. 1295, 1298 (2010) (quoting Louis Brandeis, *What Publicity Can Do*, HARPER'S WEEKLY (Dec. 1916)).

156. *See* Neil M. Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387, 389 (2008).

157. *See* BLACK'S LAW DICTIONARY 745 (6th ed. 1990).

158. *See, e.g.*, James D. Fearon, What is Identity (As We Now Use the Word)? (Nov. 3, 1999) (unpublished manuscript), *available at* https://www.stanford.edu/group/fearon-research/cgi-bin/wordpress/wp-content/uploads/2013/10/What-is-Identity-as-we-now-use-the-word-.pdf.

We want to think of identity in a third way, as "something deeper, more mysterious, and more important."[159] Psychologist Erik Erikson observed this kind of identity as "a process 'located' *in the core of the individual* and yet also *in the core of his communal culture*, a process which establishes, in fact, the identity of those two identities."[160] Julie Cohen observes, "Selfhood and social shaping are not mutually exclusive. Subjectivity, and hence selfhood, exists in the space between the experience of autonomous selfhood and the reality of social shaping."[161] Cohen goes on to assert that "[p]eople are born into networks of relationships, practices, and beliefs, and over time encounter and experiment with others, engaging in a diverse and ad hoc mix of practices that defies neat theoretical simplification."[162]

This kind of identity is the fundamental right to define who I am. This is the idea that we can define our own identities; we can say whether "I am me; I am anonymous. I am here; I am there. I am watching; I am buying. I am a supporter; I am a critic. I am voting; I am abstaining. I am for; I am against. I like; I do not like."[163] We can understand many of the protections of constitutional law in these terms—especially the political, religious, and social rights protected by the First Amendment. Indeed, our constitutional design suggests that the people, the "I am," would govern who "we are" and not the other way around.

We need to step back and see more clearly the "message" of big data to understand how it can compromise identity. Media theorist Marshall McLuhan opened his seminal 1964 book *Understanding Media: The Extensions of Man* with the oft-repeated declaration that "[i]n a culture like ours, long accustomed to splitting and dividing all things as a means of control, it is sometimes a bit of a shock to be reminded that, in operational and practical fact, the medium is the message."[164] McLuhan's maxim that "the medium is the message"[165] conveys broadly how technologies and media not only change the message but the very structure of human thought and expression. We think and act differently when we use different technologies to express ourselves or live our lives, from speaking to reading to letter writing to Google.[166] Big data technology combined with the scale

159. *See* Philip Gleason, *Identifying Identity: A Semantic History*, 69 J. AM. HIST. 910, 923 (1983).

160. *Id.* at 914.

161. Julie E. Cohen, *What Privacy Is for*, 126 HARV. L. REV. 1904, 1909 (2013).

162. *Id.* at 1910.

163. Richards & King, *supra* note 147.

164. MARSHALL MCLUHAN, UNDERSTANDING MEDIA: THE EXTENSIONS OF MAN 19 (2013).

165. *Id.*

166. *See* NICHOLAS CARR, THE SHALLOWS: WHAT THE INTERNET IS DOING TO OUR BRAINS 3 (2011).

and pace of the big metadata computer medium will change not only how we express ourselves but how we make decisions about who we are.

As citizens, we live in the early days of the marshaling of big data to help save us from terrorism and looming cyber threats. This is enabling levels of institutional surveillance of citizens (and consumers) that would previously have been technically and politically unimaginable.[167] In order to protect and serve us, institutions identify everyone. Continuous government surveillance programs aggregate minute, detailed records of our daily lives. This risks compromising our identity by stifling our intellectual privacy to think for ourselves as citizens and strengthens government power to discriminate, coerce, or selectively target critics.[168]

The commentary in the press and legal community regarding the leaks of Edward Snowden primarily focuses on breaches of privacy. Individuals understandably fear for their privacy when the government "can store such records and efficiently mine them for information years into the future."[169] Yet breaches of privacy are only part of what is at risk with big data surveillance. Individual (and national) identity now contends, for the first time, with the chilling effect of this kind of surveillance. Some will feel comforted in knowing that this surveillance exists to protect against terrorism, but others, perhaps those who find such kinds of surveillance counter to the ideals of this country, may be silenced. Moreover, as argued in the previous section, big data surveillance of this magnitude means individuals are living in a society where information shared with their service providers does not remain confidential. What will the cumulative effect on identity be from this lack of confidentiality and the specter of surveillance other than to compromise individual identity in a free society?

As consumers, our identities are increasingly being shaped by big data inferences and the companies that control them. In many regards, we want and need this control. Our identities are enlivened and protected by institutional uses of big data. Yet because they have access to substantial portions of the big metadata computer and the means and know-how to operate big data analytics, institutional power is increasing at the expense of individual identity in ways we do not yet fully understand. Institutions, often without our knowledge or consent, are collecting massive amounts of data about us which can be used and shared in secondary ways that we do not want or expect.

---

167. *See* Neil Richards, *The Dangers of Surveillance*, 126 HARV. L. REV. 1934, 1964–65 (2013).

168. *See id.* at 1935–36.

169. United States v. Jones, 132 S. Ct. 945, 955–56 (2012) (Sotomayor, J., concurring).

Since the power of big data comes from secondary uses of data sets to produce an infinite variety of insights and predictions, the more we the users use, the more government and for-profit owners of big data possess the means to use our data to influence our identity with secondary uses without our awareness. Security expert Bruce Schneier describes a feudal world where we pledge our allegiance to the companies that provide the digital devices and services we use.[170] Companies like Google, Facebook, Apple, and Amazon design and control the interfaces (TVs, iPhones, iPads, Android phones, Kindles, etc.) that consumers use and which can generate detailed histories of their every interaction. Professor Ryan Calo describes how these and other firms can employ big data to use our identities against us.[171] By applying big data analytics to our every interaction, data companies can shape consumers' identity by personalizing every part of the interaction.[172] These capabilities are "dramatically alter[ing] the capacity of firms to influence consumers at a personal level."[173]

As institutions continue to adopt big data, our identities will increasingly be shaped by institutional predictions and inferences that big data analytics allow. In many regards, we want and need this. We are enlivened by using personalized services such as Google and feel safer knowing that our identities and credit cards are protected from identity theft by financial institutions using big data analytics to detect fraud. Yet because they have access to substantial portions of the big metadata computer and the means and know-how to operate big data analytics, institutional power is increasing at the expense of individual identity in ways we do not yet fully understand. Since big data operates in legal and commercial secrecy as discussed above, the extent and nature of troubling outcomes like predicting teenage pregnancies[174] and rape victim identification[175] are just starting to be revealed, let alone understood. Given this lack of understanding, there will be certain predictions and inferences that we may want to have big data

170. *See* Michael Eisen, *When It Comes to Security, We're Back to Feudalism*, WIRED (Nov. 26, 2012), http://www.wired.com/2012/11/feudal-security/.

171. *See* Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. (forthcoming 2014).

172. *Id.*

173. *Id.*

174. *See* Charles Duhigg, *How Companies Learn Your Secrets*, N.Y. TIMES MAG. (Feb. 16, 2012), http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all (detailing Target's strategy of identifying women in their second trimester of pregnancy).

175. *See* Kashmir Hill, *Data Broker Was Selling Lists of Rape Victims, Alcoholics, and "Erectile Dysfunction Sufferers,"* FORBES (Dec. 19, 2013, 3:40 PM), http://www.forbes.com/sites/kashmirhill/2013/12/19/data-broker-was-selling-lists-of-rape-alcoholism-and-erectile-dysfunction-sufferers/.

boundaries around, and others that we will want to take off the table.

III. SECURING BIG DATA ETHICS

We need to ensure that we think ethically about big data and other new information technologies. These technologies are not "natural" and foreordained; they are the product of human choices and they will affect human values. We need to be sure that these human technologies shape the kind of society we want to have, for these technologies will shape the societies we will live in and the humans we will become.

How should we do this? As lawyers, one logical place to start would be through the creation of new legal rules. We already have many legal rules governing the processing of data. The FIPs may not be enough to protect us, but they are certainly still relevant. The FIPs have been the foundation of recent presidential, congressional, and regulatory reports studying the need to modernize privacy protection policy.[176] We have statutory schemes based on the FIPs like the Fair Credit Reporting Act ("FCRA") which was enacted to protect consumer financial information by ensuring that only the limited class of recipients with an actual need for such information could receive it, and to ensure that consumers had a meaningful opportunity to access and correct databases containing their financial information.[177] One approach to enhance privacy protections could be to expand the scope of the FCRA, which the FTC has enforced effectively for four decades.[178]

While embracing the FIPs, many propose addressing the new risks of big data by giving individuals additional control over their data. FTC Commissioner Julie Brill has called for a "Reclaim Your Name" initiative, providing for consumer protections "to reassert some control over their personal data."[179] The White House's Consumer Privacy Bill of Rights calls for a consumer right to exercise control over what personal data companies collect from

176. *See* EXEC. OFFICE OF THE PRESIDENT, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY 9 (2012), *available at* http://www.whitehouse.gov/sites/default/files/privacy-final.pdf.

177. *See* SOLOVE & SCHWARTZ, *supra* note 92, at 758–59.

178. *See* Press Release, FTC Issues Report: "Forty Years of Experience with the Fair Credit Reporting Act" (July 20 2011), *available at* http://www.ftc.gov/news-events/press-releases/2011/07/ftc-issues-report-forty-years-experience-fair-credit-reporting.

179. *See* Julie Brill, Commissioner, Fed. Trade Comm'n, Reclaim Your Name, Keynote Address at the 23rd Computers Freedom and Privacy Conference 10 (June 26, 2013), *available at* http://www.ftc.gov/sites/default/files/documents/public_statements/reclaim-your-name/130626computersfreedom.pdf.

them and how their data are used.[180] Similarly, in February 2012, the FTC issued its privacy report, "Protecting Consumer Privacy in an Era of Rapid Change," which called upon Congress to give consumers more control over their information held by data brokers.[181] With big data, however, strengthening privacy control is not enough.

Suggesting an alternative to privacy law, Professor Woodrow Hartzog makes the case for extending confidentiality law to enable a "chain-link" confidentiality regime that would contractually link the disclosure of personal obligations to protect information that moves downstream.[182] Hartzog argues that a chain of confidentiality is discoverable because we primarily access a small number of providers. The same technology that tracks us could be used to track our data flows and protect them with a "chain of confidentiality."[183] A confidentiality approach could strengthen downstream protections of data privacy and shift the focus from often hard-to-determine privacy protections. Any confidentiality regime, however, would have to be carefully tailored to not become overly restrictive and difficult to manage. One could also question the political feasibility of creating a confidentiality regime when even politically popular regimes like the National Do Not Call Registry took more than a decade to be implemented.[184]

Transparency is difficult to apply given its many paradoxes, but that should not daunt us. We need transparency to inform us of unexpected outcomes so that we can address them as they emerge. One approach could be for the FTC to call upon chief privacy officers to consider adding transparency policies to already-existing privacy policy frameworks. The adoption of transparency policies could allow companies to more freely operate while protecting consumers by allowing the FTC to bring enforcement actions when a promise of transparency is not upheld.

Whatever privacy, confidentiality, or transparency laws we develop, they should contemplate protections for metadata. Metadata offers an easier, often more relevant, and until recently, less privacy-constrained frontier for institutions to conduct surveillance. Further, the ease with which metadata can be combined with other data and the power of big data analytics allow much more information to be discerned from metadata than

180. EXEC. OFFICE OF THE PRESIDENT, *supra* note 176, at 9.

181. FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE (2012).

182. Hartzog, *supra* note 109, at 676–77; *see also* Peter A. Winn, *Confidentiality in Cyberspace: The HIPAA Privacy Rules and the Common Law*, 33 RUTGERS L.J. 617, 620 (2002).

183. Hartzog, *supra* note 109, at 678.

184. *See* Lior Jacob Strahilevitz, *Toward a Positive Theory of Privacy Law*, 126 HARV. L. REV. 2010, 2037 (2013).

dreamed of in the past. Put simply, laws need to be developed to address privacy challenges arising from the prevalence of metadata and the emerging capabilities of big data.

Additionally, given big data's power to predict and persuade us, we cannot merely have better compliance rules. There will be certain predictions and inferences that we may want to establish big data boundaries around and others that we will want to take off the table altogether. One area to consider building big data boundaries around is voting. Combined with social media, big data can shape the decision making of ourselves and others to help campaigns shape the decision they want. The 2012 Obama campaign made extensive use of big data to win the election. A large team of big data scientists and software engineers combined dozens of pieces of information on each registered voter in the United States to develop patterns to help them with fundraising and get out the vote activity.[185] While big data offers to enhance campaign fundraising activities, big data can personalize a candidate to make him appear like us and shape our voting decisions in ways that we do not yet understand. Moreover, combining big data with social networking seems to intrude upon our identity as defined by the relationships we keep and offers dangerous opportunities for incumbents to tip the scales. For example, in the most recent South Korean presidential election, it was revealed that the Korean National Intelligence Service and other state agencies posted more than 1.2 million Twitter messages to try to sway the election.[186] Utah recently passed legislation that restricts what voter data can be used for commercial purposes (e.g., data of birth).[187] The importance of the vote requires us to consider additional big data (and social media) boundaries around what campaigns, companies, and governments are allowed to do with big data analytics on voter registration records and what they are not.

Given big data's power to identify, categorize, and nudge us, we will also want to take certain big data predictions and inferences off the table. For example, in the analog world we protect the identity of rape victims. In the big data world, it was revealed that data brokers built lists of rape victims for sale.[188] We need to be ready to act to stop offensive outcomes such as this as they are revealed. We

185. *See* Sasha Issenberg, *How President Obama's Campaign Used Big Data to Rally Individual Voters*, MIT TECH. REV. (Dec. 19, 2012), http://www.technologyreview.com/featuredstory/509026/how-obamas-team-used-big-data-to-rally-voters/.

186. *See* Choe Sang-Hun, *Prosecutors Detail Attempt to Sway South Korean Election*, N.Y. TIMES (Nov. 21, 2013), http://mobile.nytimes.com/2013/11/22/world/asia/prosecutors-detail-bid-to-sway-south-korean-election.html.

187. *Voter Information Amendments*, UTAH STATE LEGISLATURE (2014), http://le.utah.gov/~2014/bills/static/sb0036.html.

188. *See* Hill, *supra* note 175.

cannot allow the use of big data algorithms, for example, to reverse engineer the return of racial-, gender-, and sex-based discrimination. We have these protections in the analog world and we will want them for the big data world as well. These will not be easy regulations to implement. They will undoubtedly get in the way of efficient decisions, but that is precisely the point. Civil rights and civil liberties are inefficient. Efficiency alone will not protect our identity.

More fundamentally, law alone is not enough to enshrine Big Data Ethics in our societies. Law has limits when things are moving quickly. Legal change is often slow and in our time of rapid technological change we are all aware that our legal rules are lagging behind our technologies. Laws we impose may cause unintended consequences of their own and unduly burden the Big Data Revolution still in its infancy. There may inevitably be a gap between active legal rules and the cutting-edge technologies that are shaping our societies and ourselves.

How should we fill this gap? We suggest that the most important way to ensure that the Big Data Revolution is a revolution that we want is to cultivate ethical sensibilities around information technologies. This can take several forms. One of them is privacy and information professionalism. Chief privacy officers, chief security officers, privacy lawyers, and data security consultants are accelerating industry norms and further institutionalizing privacy protection.[189] The International Association of Privacy Professionals ("IAPP"), the privacy industry's largest professional group, currently has more than 12,000 members—an increase of nearly 3,000 just since the beginning of 2012—which it attributes in part to the increase in the number of "Chief Privacy Officers."[190] The rapid rise of the Chief Privacy Officer offers a new seat at the table to build privacy awareness, break down organizational barriers, and enable organizations to protect privacy and prevent unexpected outcomes.

In addition to privacy professionals, other professional information ethicists have started to emerge. Google has an in-house philosopher who has argued publicly that companies should be thinking about their "moral operating system."[191] Palantir, the

189. Alec Foege, *Chief Privacy Officer Profession Grows with Big Data Field*, DATA INFORMED (Feb. 5, 2013, 1:30 PM), http://data-informed.com/chief-privacy-officer-profession-grows-with-big-data-field/.

190. *Id.*; *see also* Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy on the Books and on the Ground*, 63 STAN. L. REV. 247, 261–63 (2011); Kenneth A. Bamberger & Deirdre K. Mulligan, *Privacy in Europe: Initial Data on Governance Choices and Corporate Practices*, 81 GEO. WASH. L. REV. 1529, 1556–57 (2013).

191. Anthony Ha, *Google's In-House Philosopher: Technologists Need a "Moral Operating System,"* VENTURE BEAT (May 14, 2011, 2:47 PM), http://venturebeat.com/2011/05/14/damon-horowitz-moral-operating-system/.

rapidly growing big data innovator discussed earlier, has privacy and civil liberties engineers.[192] The President's Review Group on Intelligence and Communications Technologies recommended the creation of a privacy-and-civil-liberties policy official, to be located in both the National Security Staff and the Office of Management and Budget, and strengthened the charter of the Privacy and Civil Liberties Oversight Board.[193] Big Data Ethics needs to be part of the professional ethics of all big data professionals, whether they style themselves as data scientists or some other job description.[194]

Users have responsibility for the world that we are shaping, but in the past we have focused entirely on user choice, which is insufficient. Given the ever-increasing, ad hoc uses of big data, individuals themselves can serve as a positive feedback loop to report when bad outcomes occur. As discussed above, if institutions have transparency policies like they have privacy policies today, then users can know where to direct their concerns, and in turn the institution can quickly respond to complaints and improve sustainable uses of big data. But users alone cannot take responsibility for technologies and business practices that they do not themselves create but find themselves increasingly dependent upon.[195]

Technologists are the pioneers in this time of rapid change, and they will often see and understand big data privacy gaps before others. Technologists can lead the way to fill these gaps by rebutting "privacy is dead" beliefs and moving to advance Big Data Ethics. This is starting to happen. For example, "Privacy by Design" is a prominent set of seven information principles and best practices supported by legal scholars, regulators, and technology leaders alike.[196] The basic idea of Privacy by Design is that privacy

---

192. *See* John, *Going International with the Palantir Council of Advisors on Privacy and Civil Liberties*, PALANTIR (Jan. 29, 2014), http://www.palantir.com/2014/01/going-international-with-the-palantir-council-of-advisors-on-privacy-and-civil-liberties/.

193. *See* EXEC. OFFICE OF THE PRESIDENT, LIBERTY AND SECURITY IN A CHANGING WORLD: REPORT AND RECOMMENDATIONS OF THE PRESIDENT'S REVIEW GROUP ON INTELLIGENCE AND COMMUNICATIONS TECHNOLOGIES 35 (2013), *available at* http://www.whitehouse.gov/sites/default/files/docs/2013-12-12_rg_final_report.pdf (discussing Recommendations 26 and 27).

194. *See, e.g.*, Tam Harbert, *Big Data, Big Jobs?*, COMPUTERWORLD (Sep. 20, 2012), http://www.computerworld.com/s/article/9231445/Big_data_big_jobs_?taxonomyId=221&pageNumber=1.

195. *See* JARON LANIER, YOU ARE NOT A GADGET: A MANIFESTO 8–9 (2010) (explaining the responsibility of technologists in selecting design choices for users).

196. *E.g.*, M. Ryan Calo, *Against Notice Skepticism in Privacy (and Elsewhere)*, 87 NOTRE DAME L. REV. 1027 (2012); Deirdre Mulligan & Jennifer King, *Bridging the Gap Between Privacy and Design*, 14 U. PA. J. CONST. L. 989 (2012); Ira S. Rubenstein, *Regulating Privacy by Design*, 26 BERKELEY TECH. L.J. 1409 (2011); Peter Swire, *Social Networks, Privacy, and Freedom of*

cannot be ensured solely by regulatory oversight by government agencies; instead, effective protection of privacy also requires companies to respect the privacy of individuals by making privacy protection an ordinary but integral part of the way they do business.[197]

Technologists can also innovate to produce new technologies, business models, and best practices. A growing industry of privacy startups are starting to attract investment, such as Personal.com, which is offering personal data lockers to protect and even monetize personal data for individual benefit.[198] The Respect Network is a startup applying Privacy by Design principles to big data and establishing technology standards to support personal clouds for individuals to safely store and share personal data.[199] Jonathan Mayer and Arvind Narayanan advocate for engineers to consider the spectrum of "privacy substitutes" and quantify the trade offs between functionality and profit for consumer privacy.[200] They recommend that privacy regulators should increasingly focus on and foster available technology substitutes for privacy, not just balancing privacy risks against a growing list of countervailing societal values.[201]

Finally, big data by its very nature requires experimentation to find what it seeks. A central part of this experimentation, if we are to have privacy, confidentiality, transparency, and protect identity in a big data economy, must involve informed, principled, and collaborative experimentation with privacy subjects. To govern big data experimentation, Professor Calo proposes consumer review boards modeled on the long-standing principles of human-subject review boards created by universities to resolve ethical problems involving human-subject research.[202] Calo observes that the power relationship the experimenter and the subject require higher standards of minimizing harm or causing unfairness as a result of the experiment.[203] Given the ever increasing, ad hoc, and at times surprising secondary uses of big data, a higher standard of care

---

Association: Data Protection vs. Data Empowerment, 90 N.C. L. REV. 1371 (2012); *see also* FED. TRADE COMM'N, *supra* note 181.

197. *See generally* ANN CAVOUKIAN, PRIVACY BY DESIGN: THE 7 FOUNDATIONAL PRINCIPLES (2011), *available at* http://www.ipc.on.ca/images/resources/7foundationalprinciples.pdf.

198. *See* Joshua Brustein, *Start-ups Seek to Help Users Put a Price on Their Data*, N.Y. TIMES, Feb. 13, 2012, at B5.

199. *See generally* RESPECT NETWORK, http://www.respectnetwork.com (last visited May 6, 2014).

200. *See* Jonathan Mayer & Arvind Narayanan, *Privacy Substitutes*, 66 STAN. L. REV. ONLINE 89, 89 (2013), http://www.standordlawreview.org/artes/default/files/online/topics/66_SLR_89_MayerNarayanan.pdf.

201. *See id.*

202. *See* Calo, *supra* note 171.

203. *Id.*

model like Calo proposes would let individuals themselves serve as a positive feedback loop before bad outcomes occur.

CONCLUSION

We might well be living in the time that Licklider predicted would be "intellectually the most creative and exciting in the history of mankind."[204] Like other novel information technologies, big data presents amazing possibility to usher in a new age of discovery and innovation for mankind. We need to enable government officials to use big data to act in our defense. We want to share information with companies to let them serve us better with big data. Yet we need to think more broadly about big data so we can develop privacy ethics, norms, and legal protections to prevent important societal values like privacy, confidentiality, transparency, and identity from becoming subordinate to the new capabilities of big data.

Big data is certainly a threat to privacy, confidentiality, and identity, but it does not spell the death of law. Rules governing the way personal information flows through our society are both essential and inevitable in one form or another. But the scale of our Information Revolution means that we must think more imaginatively and broadly about what kinds of rules we want. We need to develop an approach to those rules that ensures personal information in our society flows and is used in ethical ways. This will require a social conversation that is broader than this paper. Our Big Data Revolution promises not just awareness but power—power to predict, power to shape, and power to make decisions that affect the lives of ordinary people. As in other areas of the law, sometimes good procedures will be enough, but other times we will want to put substantive limitations on what we can do with data. As we all try to harness the benefits of our new technologies without succumbing to their potential harms, developing an ethics of big data will be essential. Big Data Ethics are for everyone.

204. Licklider, *supra* note 10 at 5.