2014

Redescribing Health Privacy: The Importance of Health Policy

Frank A. Pasquale
University of Maryland Francis King Carey School of Law, fpasquale@law.umaryland.edu

Houston Journal of Health Law & Policy

REDESCRIBING HEALTH PRIVACY: THE IMPORTANCE OF INFORMATION POLICY

Frank Pasquale[1]

TABLE OF CONTENTS

[1] I wish to thank Jessica Mantel for the invitation to present this work at the University of Houston and April Moreno for great logistical assistance. Comments from participants at the roundtable were deeply appreciated.

I. INTRODUCTION

Current conversations about health information policy often tend to be based on three broad assumptions. First, many perceive a tension between regulation and innovation. We often hear that privacy regulations are keeping researchers, companies, and providers from aggregating the data they need to promote innovation.[2] Second, aggregation of fragmented data is seen as a threat to its proper regulation, creating the risk of breaches and other misuse.[3] Third, a prime directive for technicians and policymakers is to give patients ever more granular methods of control over data.[4]

This article questions and complicates those assumptions, which I deem (respectively) the Privacy Threat to Research, the Aggregation Threat to Privacy, and the Control Solution. In some contexts, strong rules regarding data acquisition, transfer, analysis, and use are key components of an innovation ecosystem. If patients will refuse to opt in to databases without strong privacy protections, then privacy is a prerequisite, not a barrier, to research and innovation.

Nor is aggregation always a threat to the types of values (ranging from due process to anti-discrimination to dignity) that privacy is meant to promote. Some types of important privacy harms can only be remedied when there is some central authority with access to all relevant databases. Fragmentation of information, far from always helping privacy, may create Kafkaesque scenarios where individuals feel helpless to correct damaging misconceptions about them.

Finally, we need to think rigorously about "control" as a be-all, end-all solution to health privacy matters. Many patients either can't be responsible (or don't want to be responsible) for exercising control over health data. Paradoxically, the sickest, most vulnerable persons may be the ones with the most data to manage—and the least time or

---

[2] I call this assumption "privacy as a threat to innovation" or "the Privacy Threat" for short.

[3] I call this assumption "aggregation as a threat to privacy" or "the Aggregation Threat" for short.

[4] I call this assumption "control as a solution to privacy concerns" or "the Control Solution" for short.

energy to take on this oft-neoliberal concept of identity management. Moreover, the more convenient access to data becomes, the more tempting ill-advised secondary uses may be. Think, for instance, of job recruiters who could "suggest" to clients that they share their health records with potential employers or "rating and ranking" firms that integrate health status into overall scores of employability and likely success. The better law and technology allow individuals to "control" their health records, the harder it is to say to anyone who asks for one, "that would be a lot of work to collect." Indeed, we already see some versions of this problem in the rise of mobile health apps, whose data protection practices are far from clear.[5] Thus, any aggressive promotion of the Control Solution must be complemented with ongoing, equally aggressive efforts to outlaw or otherwise reduce problematic uses of health data.

This article is also intended to enrich our concepts of "fragmentation" and "integration" in health care. There is a good deal of sloganeering around "firewalls" and "vertical integration" as idealized implementations of "fragmentation" and "integration" (respective). The problem, though, is that terms like these (as well as "disruption") are insufficiently normative to guide large-scale health system change. They *describe*, but they do not adequately *prescribe*.

By examining those instances where: a) regulation promotes innovation, and b) increasing (some kinds of) availability of data actually enhances security, confidentiality, and privacy protections, this article attempts to give a richer account of the ethics of fragmentation and integration in the U.S. health care system. But, it also has a darker side, highlighting the inevitable conflicts of values created in a "reputation society"[6] driven by stigmatizing social sorting systems. Personal data control may exacerbate social inequalities. Data aggregation may increase both our powers of research and our vulnerability to breach. The health data policymaking landscape of the next decade will feature a series of intractable conflicts between these important social values.

---

[5] *See* Scott Peppet, Regulating the Internet of Things, TEX. L. REV. (forthcoming, 2014), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2409074.

[6] *See* Hassan Massum & Mark Tovey, THE REPUTATION SOCIETY (2011).

II. SYSTEMATIZING THREAT SCENARIOS

The health informatics literature has developed complex typologies of threat scenarios generated by digitized health records. For example, consider this list of threats from a recent review article, *Privacy in Mobile Technology for Personal Health Care*, with respect to personal health records*:*

Identity threats: misuse of patient identities

Patients leave PHR [personal health record] credentials on public computer (identity loss)

Patients share passwords with outsiders (identity sharing)

Patients reveal passwords to outsiders (social-engineering attack)

Insiders misuse identities to obtain reimbursement (insurance fraud)

Insiders misuse identities to obtain medical services (identity theft)

Outsiders misuse identities to obtain medical services (identity theft)

Outsiders re-identifying PHI in de-identified data sets

Outsiders observe patient identity or location from communications

Access threats: unauthorized access to PHI or PHR

Patients' consent preferences, as expressed, do not match those desired

Patients' intentional (or unintentional) access beyond authorized limit

Patients' mistaken modifications, because of over-privilege or inadequate controls

Insiders' mistaken modifications, because of over-privilege or inadequate controls

Insiders' intentional unauthorized access, for curiosity or malice

Insiders' intentional modifications, to obtain reimbursement (insurance fraud)

Outsiders' intentional unauthorized access, for curiosity or malice

Outsiders' intentional modifications, for fraud or malice

Disclosure threats: unauthorized disclosure of PII and PHI

Data at rest, in the PHR

Patients' inadvertent disclosure due to malware or file-sharing tools

Insiders' inadvertent disclosure due to malware or file-sharing tools

Insiders' inadvertent disclosure due to sharing passwords

Insiders' intentional disclosure, for profit or malice

Outsiders' intentional disclosure, for profit or malice[7]

Some of these scenarios are obvious harms in their own right; others are menacing because they can lead to discrimination, improper treatment (based on inaccurate or incomplete data), or stigma (a range of unfair and/or humiliating responses to unauthorized disclosures of health problems of the data subject to family, employers, or even the public at large).

It is tempting for lawyers to try to work up a computable account of threat scenarios.[8] For many of the leading privacy theorists, a commitment to "privacy by design" would integrate legal commitments into the core technical architecture of health information technology products and services.[9] However, it is difficult to decompose extant threat scenarios in areas ranging from discrimination to unfair data practices to stigma into computable elements. Moreover, even if they were analyzable in this way in principle, it is far from certain that an engineering perspective would address deeper problems.[10]

---

[7] Avancha Sasikanth et al., *Privacy in Mobile Technology for Personal Health Care,* 45 ACM COMPUTING SURVEYS (Nov. 2012).

[8] For a definition of computability, *see* Harry Surden, *Computable Contracts*, 46 UC DAVIS L. REV. 629 (2012) ("Colloquially, the term "computable" is used when a computer can be given the means to produce a desired result (such as a mathematical computation.")).

[9] Michael Rich, *Should We Make Crime Impossible?*, HARV. J. L. & PUB. POL'Y 796, 816-18 (2013).

[10] *Cf.* Dwork & Mulligan, *It's Not Privacy, and It's Not Fair,* 66 STAN. L. REV. ONLINE 35 (2013) ("The computer science community, while acknowledging concerns about discrimination, tends to position privacy as the dominant concern. Privacy-preserving advertising schemes support the view that tracking, auctioning, and optimizing done by the many parties in the advertising ecosystem are acceptable, as long as these parties don't "know" the identity of the target. Policy proposals are similarly narrow. They include regulations requiring

The following seven examples offer narratives and scenario analysis regarding emerging threat scenarios. For example, as Nicolas Terry has shown, a threat scenario not adequately covered in the extant literature is the rise of data brokers outside the HIPAA-protected zone who can project health reputations onto identifiable individuals, even without access to records from entities covered by HIPAA.[11] The Federal Trade Commission has recently taken notice, building on the work of Terry, Scott Peppet, and other academics, to recognize the unique privacy threats posed by runaway data and a variety of other problematic scenarios. There are many other examples of diverse players making unexpected and troubling use of health information.

a. Runaway Data

Merely doing a few searches online led to a woman becoming associated with multiple sclerosis by data brokers.[12] The woman had searched online for information about a few diseases, including multiple sclerosis (MS), and subscribed to a recommendation service for physicians. That data, connected to her name via a registration form on one of the sites, was transferred to the KBM Group, a data analytics and marketing company.[13] KBM, in turn, sold it to MS Lifelines, a support network owned by two drug companies. The first time the data subject was able to even suspect what had happened was when she received promotional materials for an event for MS sufferers.

---

consent prior to tracking individuals or prior to the collection of "sensitive information," and context-specific codes respecting privacy expectations. Bridging the technical and policy arenas, the World Wide Web Consortium's draft "do-not-track" specification will allow users to signal a desire to avoid OBA. These approaches involve greater transparency. Regrettably, privacy controls and increased transparency fail to address concerns with the classifications and segmentation produced by big data analysis.").

[11] Nicolas P. Terry, *Protecting Patient Privacy in an Era of Big Data*, 81 UMKC L. REV. 385 (2012).

[12] Natasha Singer, *When Your Data Wanders to Places You've Never Been*, N.Y. TIMES, Apr. 27, 2013, http://www.nytimes.com/2013/04/28/technology/personal-data-takes-a-winding-path-into-marketers-hands.html?pagewanted=all&r_r=0..

[13] *Id.* (Those who "fill out warranty cards, enter sweepstakes, answer online surveys, agree to online privacy policies or sign up to receive e-mails from brands, they often don't realize that certain details — linked to them by name or by customer ID code — may be passed along to other companies.").

For the thick-skinned, that invitation might be shrugged off as a sort of automated gossip—ephemeral, off-base, and likely to dissipate into an ether of errant associations over time. "This is the future—get used to it," is the mantra of many Big Data firms' PR departments. But such a cavalier attitude is itself increasingly obsolete in the modern era of data science and ever-cheaper data storage.[14] Modern data science aggregates unprecedented sources (and levels) of information.[15] If the stakes are high enough, merely having a fleeting interest in a serious disease might be quite a useful variable for analysts to have. The woman associated with MS would be quite justified to worry that she might be denied life insurance, or forced to pay a higher rate, merely for a few minutes of curiosity on the Internet.[16] Moreover, when we review the hunger for health data in other contexts, even worse possibilities become evident.

b. An Unmonitorable Diversity of Data Sources

Judgments about health status do not need to be based on medical records.[17] In an era of Big Data, companies can use all manner of data to impute various medical conditions or disabilities to data subjects.[18] Consider, for instance, Charles Duhigg's report on data mining by Target: the company prides itself on knowing whether customers are pregnant.[19] Acxiom sells data about people's "online search propensity" to learn about certain ailments.[20] All this data can operate probabilistically: from a corporate perspective, what

---

[14] For storage cost estimates, *see* VIKTOR MAYER-SCHONBERGER, DELETE (Princeton, 2010).

[15] VIKTOR MAYER-SCHONBERGER, BIG DATA (2013).

[16] Individuals are pervasively, unwittingly scored on the basis of big data. PAM DIXON AND ROBERT GELLMAN, THE SCORING OF AMERICA: HOW SECRET CONSUMER SCORES THREATEN YOUR PRIVACY AND YOUR FUTURE, at http://www.worldprivacyforum.org/wp-content/uploads/2014/04/WPF_Scoring_of_America_April2014_fs.pdf (2014);

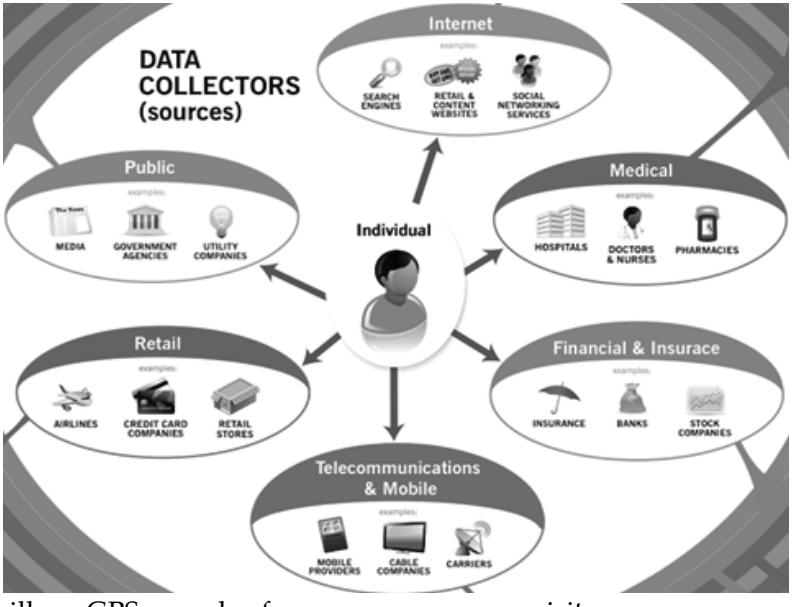
[17] Nicolas Terry, *Protecting Patient Privacy in the Age of Big Data*, 81 UMKC L. Rev. 385, 386-87 (2012).

[18] *Id.* at 386-87.

[19] 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.

[20] Lois Beckett, *Everything We Know About What Data Brokers Know About You*, PROPUBLICA, June. 13, 2014, http://www.propublica.org/article/everything-we-know-about-what-data-brokers-know-about-you.

matters is not getting any particular person's status right, but rather, playing the odds in a more informed manner. Healthcare companies are developing an interest in cognate information.[21] Cigna, for instance, has expressed an interest in fitness data.[22] Consider all the sources that could collect "health-inflected" information, such as bills



for pills or GPS records of an emergency room visit:

Image Credit: Federal Trade Commission.[23]

[21] *Id.*

[22] Jack Smith IV, *Fitbit Is Now Officially Profiting From Users' Health Data*, BETABEAT, (Apr. 18, 2014), http://betabeat.com/2014/04/fitbit-is-now-officially-profiting-from-users-health-data/#ixzz2zdt0LO2w.

[23] *Protecting Consumer Privacy in an Era of Rapid Change*, FTC, B-2 app. (Mar. 2012), *available at* http://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

And how far data brokers could go to combine and recombine those sources:

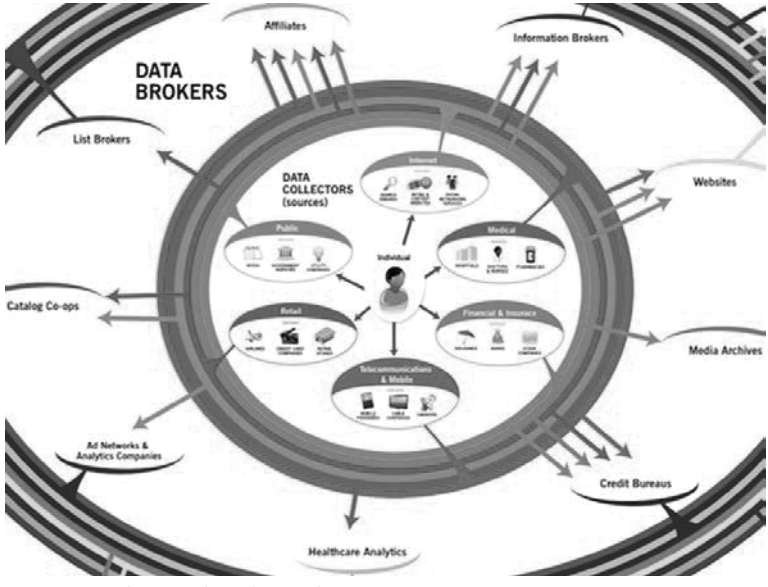


Image Credit: Federal Trade Commission.

Social networks have also intensified the surveillance of health-inflected data. These platform providers enjoy a largely deregulated online environment.[24] As social networks organized around personal health records like PatientsLikeMe, which provide novel and powerful opportunities to address health issues and to form communities, they also open the door to frightening and manipulative uses of data by firms, governments, employers, and ranking intermediaries.[25] While addressing frequently asked questions, PatientsLikeMe has stated "members should expect that every piece of information they submit (even if it is not currently displayed). . . may be shared with the community, other patients, and

[24] For a nuanced presentation of laws affecting the interaction of health professionals with social networks, *see* Nicolas Terry, *Fear of Facebook: Private Ordering of Social Media Risks Incurred by Healthcare Providers*, St. Louis U. Legal Studies Research Paper No. 2011-22 (2011). Note that the laws mentioned by Terry usually originate in the medical arena, not Internet law.

[25] ELI PARISER, THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU 3 (2011); DANIEL J. SOLOVE, THE FUTURE OF REPUTATION: GOSSIP, RUMOR, AND PRIVACY ON THE INTERNET (2007).

Partners."[26] While the company might be relied on to vet partners, its customers may have no idea about how easily information can spread. The Wall Street Journal's *What They Know* series reported that "Nielsen Co., [a] media-research firm . . . was 'scraping,' or copying, every single message off PatientsLikeMe's private online forums."[27] Health attributes connected to usernames (which, in turn, can often be linked to real identities) could have spread into numerous databases.[28] Many are not required to report to any entity on either the origin (provenance) or destination of their data.

c. Pre-PPACA Uses of Data to Deny Insurance Coverage Presage Future Uses in Other Contexts

Companies are not shy about using and distributing certain information. For those in the individual insurance market, the risk of runaway health data has already been realized. Patients who purchased antidepressants were later denied insurance repeatedly, thanks to a dossier sold to insurers. Consider, for instance, the plight of Walter and Paula Shelton, a Louisiana couple who sought insurance while in their fifties.[29] Paula had taken an antidepressant as a sleep aid, and occasionally used a blood pressure medication to relieve some swelling in her ankles.[30] Humana, a large insurer based in Kentucky, refused to insure the couple based on that prescription

---

[26] *Privacy Policy*, PATIENTSLIKEME.COM, (Mar. 5, 2012) http://www.patientslikeme.com/about/privacy.

[27] Julia Angwin & Steve Stecklow, *'Scrapers' Dig Deep for Data on Web*, WALL ST. J. (Oct. 11, 2010), http://online.wsj.com/article/SB10001424052748703358504575544381288117888.html.

[28] For connection of online and real identities, *see* Emily Steel, *A Web Pioneer Profiles Users By Name*, WALL ST. J. (Oct. 24, 2010), http://online.wsj.com/article/SB10001424052702304410504575560243259416072.html. We can glance the surface level of the extent of such aggregation by reviewing the ways sites clustered around usernames are presented in "peoplefinder" sites like Spokeo and MyLife (which sell reports to buyers interested in information associated with a name, address, phone number, or username).

[29] Chad Terhune, *They Know What's in Your Medicine Cabinet*, BLOOMBERG BUS. WK., July 22, 2008, http://www.businessweek.com/stories/2008-07-22/they-know-whats-in-your-medicine-cabinet.

[30] *Id.*

history.[31] They were not able to find insurance from other carriers, either.[32] No one had explained to them that a few prescriptions could render them uninsurable. Indeed, the model for blackballing them may still have been a gleam in an entrepreneur's eye when Mrs. Shelton obtained her drugs.

One question immediately arises: how does a company like Intelliscript get access to the data? Such pharmacy information is covered by HIPAA.[33] However, some of the insurers involved may have required applicants' permission to acquire their personal medical data to use for underwriting purposes.[34] Insurers are allowed to condition enrollment in a health plan on the "provision of an authorization requested by the health plan prior to an individual's enrollment in the health plan, if . . . [t]he authorization sought is for the health plan's eligibility or enrollment determinations relating to the individual or for its underwriting or risk rating determinations[.]"[35]

According to journalist Chad Terhune, who first reported on the Sheltons, usage of prescription data has been widespread in the individual insurance market. Insurers tailored policies to exclude

---

[31] *Id.*

[32] Patient Protection and Affordable Care Act (PPACA) 42 U.S.C §1201(4), 42 U.S.C § 2702(a)–(b)(1), 42 U.S.C. § 300gg-1(a)–(b)(1) (requiring acceptance of all applicants, but allowing limitation to certain "open or special enrollment" periods); PPACA 42 U.S.C § 1201, § 2701(a)(1)(A), 42 U.S.C.A. § 300gg(a)(1)(A) (permitting 3 to 1 age-based pricing differentials). Uninsured people like the Sheltons can count on some help from the Affordable Care Act, the landmark legislation passed in 2010. That law will require insurers to guarantee issue of policies. They can still charge people in their 50s three times as much as they charge those in their 20s, but those with a prescription history will not have to worry about flat rejections. PPACA sec. 1201(4), § 2702(a)–(b)(1), 42 U.S.C.A. § 300gg-1(a)–(b)(1) (West Supp. 1A 2010) (requiring acceptance of all applicants, but allowing limitation to certain "open or special enrollment" periods); PPACA sec. 1201, § 2701(a)(1)(A), 42 U.S.C.A. § 300gg(a)(1)(A) (permitting 3 to 1 age-based pricing differentials). They will, however, want to think about how data brokers' other forms of categorization may inform other, subtler forms of risk selection by employers and insurers.

[33] 45 C.F.R. § 160.103 (2012).

[34] For an example of such clause language, *see e.g.*, CAREFIRST HSA AGREEMENT at 36, on file with author.

[35] 45 C.F.R. § 164.508 (b)(4)(ii) (2012). Section 13405 of the Health Information Technology Act ("HITECH Act"), 42 U.S.C.A. § 17935 (d)(2)(E) (West 2012), permits insurers to pay third parties for this data, assuming the individual has authorized the disclosure.

pre-existing conditions, or to charge some members more. Companies like MedPoint and Intelliscript gathered millions of records.[36] Since 1% of patients account for over one fifth of healthcare costs, and 5% account for nearly half of costs, an insurer who can "cherry pick" the healthy and "lemon drop" the sick will be far more profitable than those who take all comers.[37] Even though the Patient Protection and Affordable Care Act's (PPACA's) guaranteed issue provisions and exchanges will help deter such underwriting practices, there are many other tactics that insurers can use to try to avoid particularly costly members. For example, "narrow networks" may be surreptitiously pushed on the vulnerable as a way of limiting insurers costs. Of course, there are federal risk adjustment and medical loss ratio rules, and state regulations that reduce the attractiveness of such strategies, but the history of health regulation reveals numerous occasions where insurers have found a way around restrictions on their underwriting practices. Moreover, there are many non-health care sectors where health information may end up being used in less regulated (and nearly undetectable) ways.

States may have wide discretion as they develop data access policies for exchanges. Insurers on exchanges will want to access data on the age and smoking history of applicants—two categories that are allowable bases for variations in premium charges. Smoking status may also be a part of wellness programs, which can lead to an up to 30% discount on insurance premiums. Although the insurer cannot use health status information to raise an individual's premiums based on PPACA,[38] the insurer could foreseeably use the information to determine single-pool risk factors related to PPACA

---

[36] Intelliscript Complaint, In the Matter of Milliman, Inc., 062-3189 F.T.C. C-4213 (2008), http://www.ftc.gov/os/caselist/0623189/080212complaint.pdf. Less harmful uses of the information may also be troubling to consumers, or may end up going beyond their original purposes. See, for instance, concerns raised in *Weld v. CVS Pharmacy* (A pharmacy sold names and contact information of customers to allow a direct marketer to target customers with specific medical conditions). Weld v. CVS Pharmacy, Inc., No. CIV. A. 98-0897F, 1999 WL 494114, at *1 (Mass. Super. Ct. June 29, 1999).

[37] William W. Yu and Trena M. Ezzati-Rice, *Concentration of Health Care Expenses in the U.S. Civilian Noninstitutionalized Population*, AGENCY FOR HEALTHCARE RESEARCH AND QUALITY (2005), http://www.meps.ahrq.gov/mepsweb/data_files/publications/st81/stat81.shtml.

[38] 42 U.S.C. § 1201.

or overall plan premiums.[39]

Indeed, the ACA may strengthen insurers' hands. In addition to the risk determination, an insurer could, for example, seek access to pharmaceutical records to determine if someone who professes to be a non-smoker has taken smoking-cessation medications in the recent past. Because the individual is protected from denial of coverage or premium increases based on most health status issues, the insurer could argue that there is even less reason to worry about the coercive nature of requiring the authorization. They are simply looking to determine proper eligibility and prevent fraud.

d. Data Laundering: How Anti-Discrimination Rules in Health Care Could Push the "Sickness-Avoidance" Strategy to the Hiring Stage, and into Credit Decisions

Even if regulators check insurers' efforts to avoid high-risk individuals (or shift them to undesirable plans), employers may adopt pretextual tactics to drive them away as employees. Given the pervasiveness of self-funded plans, this is a serious concern. Data-driven managers will want to avoid the productivity drag of a sick or otherwise impaired workforce.[40]

Moreover, don't expect these methods to be easy to detect. Where exactly is the line to be drawn between characterizing a potential employee as 1) diabetic, 2) in a "diabetic-focused household" (to use a category publicly disclosed by a data broker), 3) concerned about diabetes, 4) having a demanding home life (a determination that may in part be based on proprietary formulas extrapolating the effect of the data that would lead to attributions 1, 2, and 3)? Any effort to expand the protected zone beyond 1 is likely to draw resistance from businesses enamored with "big data" methods of increasing productivity, particularly because it would likely require extensive auditing of business records. But a narrow focus on explicit use of 1 creates only a patina of privacy, giving people a false sense of security about limits on the role of health

---

[39] 42 U.S.C. §§ 1341–43.

[40] For a description of such trends drawing on research on data borker practices see FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 26 (2015).

status as a determinant of their fortunes. Without extending zones of protection well beyond HIPAA-covered entities, health privacy law risks mirroring the "security theater" that plagues homeland security operations.[41]

While the intricate details of the Omnibus HIPAA rule are specified and litigated,[42] data brokers continue gathering information, and making predictions based on it, entirely outside the HIPPA-protected zone.[43] It is increasingly difficult for those affected to understand (let alone prove) how health-inflected data affected decision-making about them, but we now know that health-based scoring models are common.[44] While the sheer amount of data gathered by reputational intermediaries is immense, the inferences they enable are even more staggering. Unattributed data sources are available to be pervasively deployed to make (or rationalize) critical judgments about individuals. Even if some of those judgments violate the law, there is no systematic auditing of data used by large employers in their decision-making, and there are ample pretexts to mask suspect or illegal behavior.

Given the spread of health-inflected data, there are bound to be discriminatory uses of information either not covered by extant privacy or anti-discrimination laws or undetectable by workers.[45]

---

[41] "Security theater refers to security measures that make people feel more secure without doing anything to actually improve their security." Bruce Schneier, *Beyond Security Theater*, SCHNEIER ON SECURITY (June 9, 2014 4:42 PM) https://www.schneier.com/blog/archives/2009/11/beyond_security.html; JOHN MUELLER, Overblown: HOW POLITICIANS AND THE TERRORISM INDUSTRY INFLATE NATIONAL SECURITY THREATS, AND WHY WE BELIEVE THEM (2006).

[42] *See* Plaintiff's Memo in Support of its Motion for a Preliminary Injunction, Adheris, Inc. v. Dep't of Health and Human Servs., No. 1:13-cv-01342-EGS (D.C. Jan. 2010).

[43] Sarah Ludington, *Reining in the Data Traders: A Tort for the Misuse of Personal Information*, 66 MD. L. REV. 140, 183 (2006). There are also legitimate worries about discriminatory uses of information either not covered by extant privacy or anti-discrimination laws, or undetectable by workers. *See* Hoffman, *supra* note 41.

[44] Danielle Citron and Frank Pasquale, *The Scored Society*, 89 WASH. L. REV. 1, 4 (2014); Pam Dixon & Robert Gellman, *The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future*, WORLD PRIVACY FORUM, Apr. 2, 2014 at 62 (describing health scores).

[45] Sharona Hoffman, Employing E-Health: The Impact of Electronic Health Records on the Workplace, 19 KAN. J. L. & PUB. POL'Y 409,410-11(2010).

Efforts to assure the fairness and accuracy of such reputation-affecting information have not caught up to technological advances in producing it. For example, an investigating office may tailor its software to assure that the most damaging information available about a person (from its perspective) comes up first in whatever databases it queries. The applicant would need to use the same personalizing software to be fully aware of all the negative information such a search was generating. Yet, trade secrecy and contracts will likely prevent him from ever accessing an exact replica of the programs used by the educators, employers, landlords, bankers, and others making vital decisions about his future. Some digital scarlet letter could be floating in the ether, prominent to those with certain filtering programs, and virtually invisible to the person it stigmatizes. Health-inflected information from entities not covered under HIPAA can be a critical source of correlations, profiles, and attributions.

Such profiles have already started to enter into the world of credit cards. Lenders are moving beyond credit scoring to "credit analytics," which tracks a consumer's every transaction and adjusts terms of credit accordingly. Internet ad targeting for credit is even less regulated. A major retailer like Target can "know" a customer is pregnant before even other family members do, simply by crunching the numbers on a sufficiently large data set of purchases and doing pattern recognition.[46] Not surprisingly, some customers have found it creepy to start receiving pregnancy-related ads simply on the basis of

[46] Julie Brill, Comm'r Fed. Trade Comm'n, Big Data, Big Issues at Fordham University School of Law (March 2, 2013) available at http://www.ftc.gov/publuc-statements/2012/03/big-data-bigissues (recounting Charles Duhigg Article on Target); *see also*Andrew Couts, *What's the NSA Picking Out of Your Phone Calls? Just 'Unvolunteered Truths*,' DIGITAL TRENDS (Aug. 31, 2013),http://www.digitaltrends.com/mobile/whats-the-nsa-picking-out-of-your-phone-calls-just-unvolunteered-truths/#ixzz2djyaJ5LL ("The public may have first become aware of [such trends] 2012, thanks to an article in The New York Times Magazine by Charles Duhigg. You may remember its bombshell anecdote: A father, having discovered coupons from Target offering discounts on baby gear sent to his young daughter, became outraged at the big box retailer for apparently trying to coerce the teenager into having sex. In fact, Target's in-house data analysts had devised an algorithm that deduced whether particular customers were pregnant based on seemingly random changes in their purchasing habits."); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE HIDDEN ALGORITHMS BEHIND MONEY AND INFORMATION 35 (2014) (describing the pattern recognition process in greater detail).

big data analysis of their purchases of large bags, bright blue rugs, and zinc supplements. Target responded, not by explaining to customers how it came to its conclusions, but by mixing more non-pregnancy-related ads into the circulars targeting expectant mothers. Target also does not appear particularly adept at protecting its data trove; in 2013, hackers stole "mailing and email addresses, phone numbers or names, [and] the kind of data routinely collected from Target customers during interactions like shopping online."[47] We do not know how that information will circulate. This emerging threat scenario is one key reason for requiring all data brokers, and those using their services, to document exactly where they get and where they send their data. Otherwise, data from breaches could easily be laundered.[48]

Credit card companies have not ignored the lesson; indeed, they may have pioneered it. As the *New York Times* reported, some "companies started cutting cardholders' credit lines when charges appeared for . . . marriage therapy because data indicated [it was a sign] of desperation or depression that might lead to job loss."[49] What about people who pay cash for such counseling (a right guaranteed by HITECH, though inadequately implemented at present)? Data broker Acxiom appears to be using proxy data for health status when it markets "intimate details like whether a person is a 'potential inheritor' or an 'adult with senior parent,' or whether a household has a 'diabetic focus' or 'senior needs,'" to corporate clients.[50] One wonders to what extent Acxiom or other data brokers use information like Amazon sales or web browsing details to segment populations in this way. Journalist Emily Steel has reported an

---

[47] Elizabeth A. Harris & Nicole Perlroth, "*For Target, the Breach Numbers Grow*, N.Y. TIMES (Jan. 10, 2014), *available at* http://www.nytimes.com/2014/01/11/business/target-breach-affected-70-million-customers.html?_r=ai.

[48] Frank Pasquale, *Good Fences Make Better Data Brokers*, 46 GA. L. REV. 657, 659 (2012) available at http://cyber.jotwell.com/good-fences-make-better-data-brokers/.

[49] Charles Duhigg, *What Does Your Credit Card Company Know About You?* N.Y. TIMES (May 12, 2009), http://www.nytimes.com/2009/05/17/magazine/17credit-t.html?_r=0&pagewanted=print.

[50] Natasha Singer, *When Your Data Wanders to Places You've Never Been*, N.Y. TIMES (April 27, 2013), http://www.nytimes.com/2013/04/28/technology/personal-data-takes-a-winding-path-into-marketers-hands.html?pagewanted=all&r_r=0.

uptick in integration of various sources of information among data brokers.[51] Facebook, for instance, appears intent on integrating on and off-line data. On the other hand, a revelation of the full details may never happen, given the importance of trade secrecy in the industry.[52]

e. False Accusations of Substance Abuse/Addiction

Data brokers are keen to monetize their information trove—but accuracy does not always correlate with profit-maximization. As the pace of deal-making picks up, more individuals are affected by inaccurate, incomplete, or wrong information. For example, ChoicePoint is a data broker that maintains files on nearly all Americans. It mistakenly reported a criminal charge of intent to sell and manufacture methamphetamines in Arkansas resident Catherine Taylor's file.[53]ChoicePoint corrected the information when notified about the error, but other companies that had bought Taylor's file from ChoicePoint did not automatically follow suit.[54] The free-floating lie ensured rapid rejections of her job applications, and she could not even obtain credit to buy a dishwasher.[55] Some companies corrected their reports in a timely manner, but Taylor had to nag others repeatedly, and even took one to court.[56]

Taylor found that her effort to correct all of the erroneous meth conviction entries was overwhelming and emotionally taxing. "I can't be the watchdog all the time," she told a *Washington Post* reporter.[57] It

---

[51] The Leonard Lopate Show, *Privacy and Big Data (2003), available at* http://www.wnyc.org/shows/lopate/2013/jun/18/privacy-and-big-data/ (based on series of FT articles, particularly focusing on integration of Facebook data into files of real-space brokers).

[52] Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. 41, 42 (2013) ("While big data pervasively collects all manner of private information, the operations of big data itself are almost entirely shrouded in legal and commercial secrecy. We call this the Transparency Paradox.").

[53] Ylan Q. Mui, *Little-Known Firms Tracking Data Used in Credit Scores*, WASH. POST (May 21, 2011), http://www.washingtonpost.com/business/economy/little-known-firms-tracking-data-used-in-credit-scores/2011/05/24/gIQAXHcWII_print.html.

[54] *Id.*

[55] *Id.*

[56] *Id.*

[57] *Id.*

took her four years to find a job, even after the error was uncovered, and she was still rejected for an apartment.[58] Taylor ended up living in a house purchased by her sister, and claimed the stress of the wrongful accusation exacerbated her diabetes and heart problems.[59]As Elizabeth DeArmond has observed, the "power of mismatched information . . . to disrupt or even paralyze the lives of individuals has grown dramatically."[60] For every Catherine Taylor, actually aware of the data defaming her, there may be thousands of other victims entirely unaware of dubious scarlet letters besmirching their digital dossiers.

Economic theory suggests that there is little market incentive for dominant data firms or their customers to invest much time in rehabilitating those harmed by unfair data practices. As Joel Waldfogel has documented, market forces can systematically erode the availability of "niche" outputs; it may make little sense for a pharmaceutical firm to invest in equipment to make orphan drugs if there are under 200,000 customers who want to buy them (absent extant government efforts directly designed to incentivize the creation of orphan drugs).[61] "Niche" inputs such as stigmatized laborers or borrowers may face a similar fate, especially in an era of mass un- and under-employment, if government fails to demand or incentivize opportunities for their reputational rehabilitation when it is warranted.

The secrecy of data mining raises Kafkaesque possibilities. If the records of firms like ChoicePoint were more subject to audit, it would be far easier to stem the spread of dubious runaway data. Instead, much of what they do is hidden as "proprietary methods." Consider what might happen to an unfairly maligned individual if he clears his name in multiple databases by hiring an attorney. What if there are

---

[58] Ylan Q. Mui, *Little-Known Firms Tracking Data Used in Credit Scores*, WASH. POST, May 21, 2011, http://www.washingtonpost.com/business/economy/little-known-firms-tracking-data-used-in-credit-scores/2011/05/24/gIQAXHcWII_print.html.

[59] *Id.*

[60] Elizabeth D. De Armond, *Frothy Chaos: Modern Data Warehousing and Old-Fashioned Defamation*, 41 VALPARAISO UNIV. L. REV.1061, 1061 (2007).

[61] Joel Waldfogel, *The Tyranny of the Market* (Harvard Univ. Press, 2007). I choose the number 200,000 to refer to the criteria of federal incentive programs to promote orphan drug development. We may need to consider helping "reputational orphans" as well.

variables in the "black box" scoring process penalizing "litigiousness"? Consumers may be digitally stigmatized merely for trying to stop their stigmatization. Who wants to do business with a person inclined to sue? Merely trying to contest one's score could reduce it. For those with a broader goal of the human good than maximum convenience for business, the lack of due process in such systems is deeply troubling.

f. Positive Information: The Rise of the Personal Prospectus

Consider the following hypothetical: Joe Smith is a relatively healthy guy on the cusp of middle age.[62] He switches jobs, and applies for life insurance at his new employer. The insurer requires a visit from a "Portamedic," who draws blood, takes a urine specimen, measures Joe's height and weight, and takes his blood pressure. "108 over 68—a little low," the visiting doctor remarks, and enters it in the chart. "I think that's because I exercise every day for an hour," Joe protests. "Do you have a gym membership?" the doctor replies. "No, I run outdoors," Joe says. "Do you keep track of that with a fitness app?" the tech-savvy doctor asks. Again, Joe demurs. The doctor writes down nothing. A few weeks later, the life insurer turns down Joe's application for coverage.

What's the lesson here? Should Joe have bought a "FitBit" to append to his arm while running? Or perhaps carried his cellphone and subpoenaed his carrier (or the NSA) for the location information it constantly emits? Maybe there is no particular lesson. Joe has no idea why the insurer turned him down, and trade secrecy laws protect the insurer's prerogatives to secrecy. But, people like Joe will increasingly be seeking to document their health and good habits for life insurers, wellness programs, and even potential future spouses.

The Presidential Commission for the Study of Bioethical Issues issued a report titled *Privacy and Progress in Whole Genome Sequencing* in 2012, which brought up some of the novel threat scenarios involved in probabilistic analyses of genomic information:

> [I]n many states someone could legally pick up a discarded coffee cup and send a saliva sample to a commercial sequencing entity in an attempt to discover an individual's predisposition to

[62] This example is drawn from someone known to the author.

> neurodegenerative disease. The information might then be misused, for example, by a contentious spouse as evidence of unfitness to parent in a custody case. Or, the information might be publicized by a malicious stranger or acquaintance without the individual's knowledge or consent in a social networking space, which could adversely affect that individual's chance of finding a spouse, achieving standing in a community, or pursuing a desired career path.[63]

On the other hand, it should be clear from Joe Smith's story that each of these threat scenarios is also an opportunity scenario. Individuals may want to tout their *lack* of predisposition to neurodegenerative disease, or any of a number of other conditions probabilistically indicated by certain data.

For each person "trashed" by the new health data systems, there are many others angling for a competitive edge by proving how healthy they are. Although the "wellness programs" revitalized by the Patient Protection and Affordable Care Act have some strict guidelines about use of biomarkers, hunger for positive health data is growing. Reputational reporting has gone far beyond the three major credit bureaus, which have a well-established system for giving consumers their files. Smaller firms maintain largely unknown databases about nearly every aspect of individuals' lives, including health-inflected data.[64] Alliant Data, for instance, documents missed monthly payments for gym memberships. Apps like FitBit and LoseIt are high-tech diaries of individual exercise and eating habits.

Social sorting is big business. Bosses and bankers crave predictive analytics that will decide who will be the best workers and borrowers—and health status can heavily influence one's work life and financial well-being. A high percentage of US bankruptcies are medical in nature.[65] As "Big Data" enchants managers, our economy

---

[63] PRESIDENTIAL COMM'N FOR THE STUDY OF BIOETHICAL ISSUES, *Privacy and Progress in Whole Genome Sequencing* 2 (2012).

[64] *See* Mui, *supra* note 58. Other examples include: ChexSystmes and TeleCheck track bounded checks; payday lenders report "deadbeats"toTeletrack. The National Communications, Telecom and Utilities Exchange uses data from several large companies to set recommended deposits for cable andutility subscribers, but won't reveal who they are.

[65] NPAF, *Medical Debt, Medical Bankruptcy and the Impact on Patients* (2012),available at http://www.npaf.org/files/Medical%20Debt%20White%20Paper%20Final_0.pdf ("Medical expenses are contributing factors in over 62% of individual (as opposed to business) bankruptcy filings, and the number of bankruptcy filings attributing medical debt as a significant factor is increasing.).

rewards startups that can identify future big spenders and bankrupts, using whatever data is at hand. Consumers can play the game, too, parleying physical hardihood into financial fitness. But, this opportunity is double-edged—as much as it promotes the healthy, it will leave behind the sick. The trend also takes procyclical dimensions of our macro-economy to the personal realm, expanding opportunities for people while they are healthy and gaining wealth, and suddenly shutting them out when they become sick and most need an extra boost.

Individuals are increasingly volunteering information about themselves in order to stand out from the crowd. Consider an HR office's choices, as applicants start to put college grade point averages (GPAs) on their resumes. At some point, a critical mass of applicants may include a GPA, and as that happens, those who don't include it may look like they are trying to hide their grades.

When such self-disclosure reaches a critical mass, a tipping point is reached, and everyone essentially *must* disclose in order to avoid being stigmatized as someone with something to hide. Economists of information label this process "unraveling." As "rapidly changing information technologies are making possible the low-cost sharing of verified personal information for economic reward," the ultimate effect will be little different than if snooping employers, government officials, and other decision-makers could directly *demand* damaging information.[66] The "economics of signaling" is thus a critical consideration in the field of privacy governance.

Promoting one's good health may seem like a vaguely creepy or dishonorable competitive tactic now. But Americans' standard response to new surveillance technologies is a flurry of anxiety, rapidly followed by acquiescence. Journalist Kashmir Hill aptly deemed these short-lived moments of resistance "privacy freakouts," as ephemeral as a brush fire. Meanwhile, the firm MyBodyCount is already marketing a "health score" designed to permit individuals to use verified optimal biomarkers as a "currency" to leverage discounts

[66] Anita Allen, *Dredging Up the Past: Lifelogging, Memory, and Surveillance*, 75 U. CHI. L. REV.52, (2008); Emily Singer, *The Measured Life*, MIT TECH. REV. (2011), *available at* http://www.technologyreview.com/.

or other advantages.[67]

At some point, the instinct for privacy may not even serve as a speed bump for the monitoring juggernaut. Thousands of small games and rewards will give the emerging panopticon a human face. Carnegie Mellon professor Jesse Schell gave the following example at a Silicon Valley confab, explaining the fun and games possible in a pervasively monitored world:

> Well, I think it'll be like this: You get up in the morning to brush your teeth and the toothbrush can sense that you're brushing your teeth, and so, hey, good job for you! Ten points for brushing your teeth. And it can measure how long, and you're supposed to brush 'em for three minutes, and you did. And so you get a bonus for that. Hey, you brushed your teeth every day this week, another bonus. All right, and who cares? The toothpaste company, the toothbrush company; the more you brush, the more toothpaste you use. They have a vested financial interest.[68]

So too do corporate leaders in charge of wellness programs. Employers will want to know if workers are getting regular exercise, eating well, and flossing, and can discount up to 30% of health insurance premiums for well-behaved health maximizers. Data-driven decisionmakers want a "fitter, happier, more productive" labor force (in the immortal words of Radiohead).[69]

Schell describes a larger social process of "gamification": the sequencing of tasks into clear intervals with rules, scores, and immediate feedback.[70] The premise is hopeful. 'Gamify' toothbrushing, and there will be fewer cavities. 'Gamify' driving (say, by giving insurance discounts for following the speed limit), and there will be fewer accidents. Sequence the tasks of health

---

[67] Press release, *MyBodyCount Launches First Universal Health Score Based on Lifestyle Risk Factors*, MYBODYCOUNT (June 24, 2013) at http://mybodycount.com/wp-content/uploads/2013/10/mbc_launch.pdf

[68] Jesse Schell, *When Games Invade Real Life*, beginning at 21:14-21:46 on TED Talk (2010),http://www.ted.com/talks/jesse_schell_when_games_invade_real_life.html.

[69] RADIOHEAD, *Fitter Happier*, *on* OK COMPUTER (Capitol Records 1997).

[70] *See generally* PHILLIP MANNING, ERVING GOFFMAN AND MODERN SOCIOLOGY (1992) (noting Erving Goffman's views on social life as a game); *See also* JANE MCGONIGAL, REALITY IS BROKEN: WHY GAMES MAKE US BETTER AND HOW THEY CAN CHANGE THE WORLD (2011); *see also* David Golumbia, *Game of Drones*, *available at* https://www.academia.edu/2260255/Game_of_Drones.

maintenance into a Pavlovian reward structure, and a once-intimidating mass of hard-to-remember suggestions and guidelines seems more manageable. The games aim to enliven ordinary life, while subjecting citizens to a level of scrutiny once reserved for public figures. But they also create a real danger of all-pervasive modulation of daily life, according to agendas that may be unknown by consumers (and may well diverge from their own interests, ranging from autonomy to privacy).

The Federal Trade Commission has expressed some interest in understanding what data brokers are doing, but it is technologically outmatched. Consumer protection agencies have nowhere near the staff they would need to monitor all companies trafficking in reputational data.[71] Moreover, the agency's extant fines for privacy violations are so puny that they would deter few data brokers.

III. RECONSIDERING THE PRIVACY THREAT, THE AGGREGATION THREAT, AND THE CONTROL SOLUTION IN LIGHT OF CONTEMPORARY DATA FLOWS

Now that a range of threat scenarios is on the table, let's consider the types of issues they raise more systematically. They are displayed in abbreviated fashion in the chart below:

Baseline Scenarios

	A) Some Principal Actors Involved	**B) Info Status**
1) Runaway Data (probabilistic assessment leads to attribution of disease)	Online Info Sources Marketing Firm	Attribution: False Probabilistic assessment: "True" Helpfulness or harmfulness depends on use
2) Mystery Data (thanks to tough TOS)	Social Network Managers (use of boilerplate terms to eviscerate	Unknown

[71] Peter Maass, *Your Privacy Watchdog: Low-Tech, Defensive, Toothless*, WIRED at 4 (June 28, 2012, 6:30 AM), http://www.wired.com/2012/06/ftc-fail/.

	expectation of privacy)	
3) Denial of Insurance Using Health Data Outside "HIPAA Zone" (via legislation or consent)	Insurers MedPoint, Intelliscript	Use of drug (here, Prozac) true Implication (depression) false
4) Employment and Credit Discrimination via Health Data Laundering	Employers Health-Inflected Data Brokers	Unknown
5) False Accusations of Substance Abuse/Addiction (Catherine Taylor example)	Data Brokers Clients of Data Brokers	False and Harmful
6) Positive Information: The Personal Prospectus (gym membership; apps)	Third Party Apps Independent Verification Entities	True and Helpful
7) Credit Discrimination via Health Reputation or Data Laundering	Credit Card Companies Predictive Analytics Firms	Opinion; could be either helpful or harmful

Recall the Aggregation Threat assumption: that the combination of fragmented data is seen as a threat to its proper regulation, creating the risk of breaches and other misuse. Some of our examples support it. For example, [3], [4], and [7] all rely on the advanced databases that combine disparate sources of information. But consider [5], Catherine Taylor's battles with data brokers over a false meth accusation. Her problems arose primarily because data was fragmented. Corrections in ChoicePoint's database did not automatically populate other databases. Similarly, the "runaway data" in [1] and mystery data in [2] is as much a function of fragmentation as of integration. A more centralized health data infrastructure might be more amenable to monitoring, segmentation, regulation, and breach detection.

What about the "Control Solution?" Those aiming to perfect their

"personal prospectuses" with positive information (vis-a-vis example [7]) may be enthusiastic about "personal data vaults,"[72] the "Data Map,"[73] and similar innovations. But what about those who would rather not share? The practical obscurity now afforded by data fragmentation may help them resist "requests" from powerful parties that they make their data selves completely legible. Thus, when we hear that "personal data stores (PDS), 'infomediaries,' Vendor Relationship Management (VRM) systems, and federated and distributed social networks" are "solutions" to a privacy problem, we should always ask in response: *whose* privacy problem?[74]

Some experts argue that data vaults will help inculcate an ethic of self-care known as "cyberhygiene," encouraging people to be more careful about exactly when and how they disclose information. But, the "privacy hygiene" recommendations of digital gurus are often rendered obsolete by advances in technological countermeasures.[75] Think your "good browser hygiene" renders your web surfing less traceable? Welcome to the world of "device fingerprinters" that tie anything you view on the Internet on a device you own to a persistent identifier.[76] In the wake of the PRISM revelations in June 2013, tech-savvy journalist Timothy B. Lee advised his readers to use Tor to "browse anonymously."[77] In August, security expert Bruce Schneier raised the possibility that "Tor has been compromised."[78] The HHS's "Wall of Shame" testifies to the extent and severity of

---

[72] Jerry Kang et al., *Self-Surveillance Privacy*, 97 IOWA L. REV. 809, 812, 829 (2012).

[73] Latanya Sweeney, the DataMap, www.thedatamap.org (last visited Feb. 26, 2014).

[74] Arvind Narayanan et al., *A Critical Look at Decentralized Personal Data Architectures* (Feb. 20, 2012), *available at* http://randomwalker.info/publications/critical-look-at-decentralization-v1.pdf.

[75] *Cf.* Jathan Sadowski, *Why We Should Wash Our Hands of "Cyber-Hygiene*, FUTURE TENSE (June 19, 2013) (discussing the destructive impact of the "privacy hygiene" rhetoric).

[76] Julia Angwin & Jennifer Valention-Devries, *Race is on to 'Fingerprint' PCs, Phones*, WALL ST. J. Nov. 30, 2010, http://online.wsj.com/article/SB10001424052748704679204575646704100959546.html.

[77] Timothy B. Lee, *Five Ways to Stop the NSA from Spying on You*, WASH. POST, June 10, 2013 12:22 PM, http://www.washingtonpost.com/blogs/wonkblog/wp/2013/06/10/five-ways-to-stop-the-nsa-from-spying-on-you/.

[78] Bruce Schneier, *Has Tor Been Compromised?*, SCHNEIER ON SECURITY (Aug. 6, 2013 1:42 PM), http://www.schneier.com/blog/archives/2013/08/has_tor_been_co.html.

unauthorized disclosures of even highly regulated health data.[79]

A growing technical literature has pointed out the flaws in "Control Solutions." As Narayanan et al. argue:

> [M]ore control over personal data almost inevitably translates to more decisions, which leads to cognitive overload. . . . [S]ince users lack expertise in software configuration, security vulnerabilities may result. A related point is that users may be unable to meaningfully verify privacy guarantees provided through cryptography.[80]

We should also note that "control solutions" may require significant alteration of the ways in which data collectors maintain their databases. We can only solve problems like "runaway data" for medical reputations if data controllers are required to attach an identifiable symbol (such as "+H+" for health) into metadata for observations recording or predicting health information. That would allow the data controllers to filter out health information in reports and calculations performed in response to sensitive queries in the employment and insurance contexts. In other words, we need to develop a basic infrastructure of annotation and filtering before we can begin to prevent data misuse.[81]

While Narayanan and his coauthors expertly analyze the "control solutions" failures for their own users, we should also be concerned with an issue suggested by scenarios [6] (positive information) and [7] (credit discrimination): namely, whether those who don't manage to maintain good "data hygiene" should be "easy pickings" for various sorting and targeting programs.[82]

The spread and use of data is extraordinarily complex. Consider, for instance, just one depiction of the types of information exchange

---

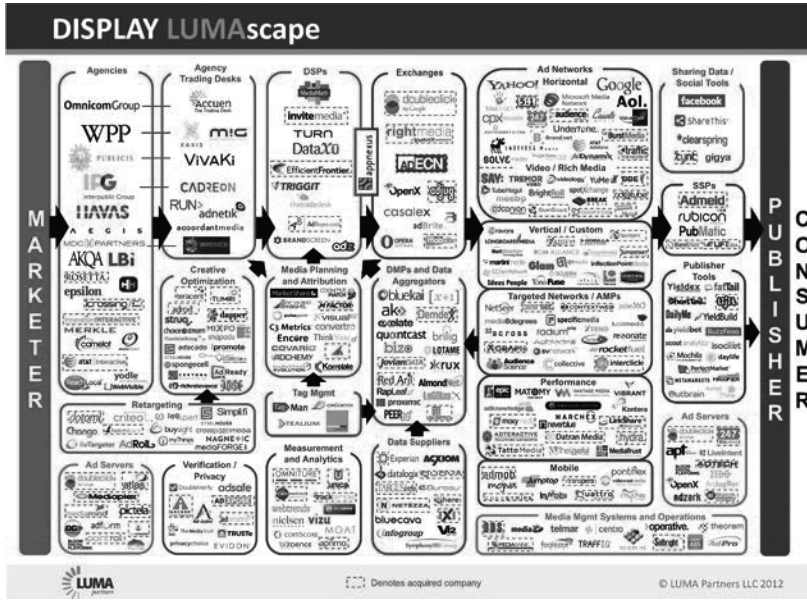
[79] *See* U.S. DEP'T OF HEALTH & HUMAN SERVCES, BREACHES AFFECTING 500 OR MORE INDIVIDUALS, http://www.hhs.gov/ocr/privacy/hipaa/administrative/breachnotificationrule/breachtool.html (last visited Apr. 3, 2014).

[80] *See* Narayanan et al., *supra* note 74.

[81] Perhaps E-discovery and deduplication software may be tweaked to remove health data without the need of annotation. Computer science researchers are already investigating similar projects, such as removing social security numbers and certain names from court records due to be digitized and made available online. But I believe such innovation could be a long time in coming.

[82] *Cf.* Sadowski, *supra* note 75.

now occurring online between advertisers (including those advertising credit offers), publishing platforms, and the multiple exchanges that stand between them:



This slide was prepared by a venture capitalist,[83] one of those rare individuals with a strong enough interest in the matter to actually map the flow of data. What would a consumer do with a report that indicated the types of information each of these entities processed about her as they combined to serve up offers of credit on the Internet? Consider just the "Data Suppliers" and "Ad Networks" boxes—there are numerous firms.

At its worst, the Control Solution deconstructs the social value of privacy into a series of individual *opportunities*—preferentially available to the wealthy, healthy, savvy, and less time constrained—to assure that their digital selves "look better" than those of others (i.e., the people without the time to set up personal data stores, transact with infomediaries, hire reputation managers, carefully

[83] Terence Kawaja, *Display LUMAscape*, LUMA PARTNERS, http://www.lumapartners.com/lumascapes/display-ad-tech-lumascape (last visited Mar. 23, 2014).

review Terms of Service, and install encryption on their computing systems). The digerati may develop optimized, secure Personal Health Records. But unless that transition is carefully planned, it could leave the rest of us as soft targets for a profit-hungry surveillance apparatus eager to sell profiles to the highest bidders. If new Control Solutions are exercised only (or even primarily) by the already-advantaged, they may merely help an upper crust of society leverage monetary advantage into information advantage.[84]

It is debatable whether "privacy as self-protection via shrewd data disclosure" is a self-concept that policymakers should seek to cultivate. In a world where consumers are expected to zealously guard their data (or suffer the consequences), consumers most in need of fair information practices are least likely to have the resources to actually demand and secure their data.[85] The proper allocation of surveillance has very little relationship with users' desire to pay for privacy, and indeed may be inversely correlated with it.

At their worst, tactics of information control may become one more set of rules that the haves manipulate to increase their advantages over the have-nots. In a world where persons are persistently ranked and stigmatized via data collection, an equilibrium featuring wealthy individuals who have purchased privacy and poorer individuals who cannot afford it may be worse than equilibrium where no one has access to this product. As data scientist Cathy O'Neil observes:

> There are very real problems in the information-gathering space, and we need to address them, but one of the most important issues is that the very people who can't afford to pay for their reputation to be kept clean are the real victims of the system. . . . [T]hrough using the services from companies Reputation.com and because of the nature of the personalization of internet usage, the very legislators who need to act on behalf of their most vulnerable citizens won't even see the

---

[84] *Cf.* Pierre Bourdieu, *The Forms of Capital, in* HANDBOOK OF THEORY AND RESEARCH FOR THE SOCIOLOGY OF EDUCATION 46, 46-47 (J. Richardson ed., 1986) (on translation of monetary into social capital via education system); Amanda MacBlane, *In Conversation with Francie Ostrower*, NEW MUSIC BOX (March 1, 2003).

[85] Michele Estrin Gilman, *The Class Differential in Privacy Law*, 77 BROOK. L. REV. 1389, 1423 (2012).

problem since they don't share it.[86]

Perhaps one day services like Reputation.com will scale to offer more affordable "products," to a mass audience, but even this market-based model would fail because privacy protection is not remotely a "thing." Rather, privacy is a social practice.[87] One can almost never contract for a certain level of privacy protection and expect for that mere assurance to be the end of the matter. In a world of constantly evolving threats and vulnerabilities, restricting data flows can be as complex and beset by asymmetric information—and uncertain outcomes—as health care. Users have so many points of vulnerability that it seems futile to focus on fixing any one of them.

For example, a consumer could refrain from talking about personal illnesses on Gmail or Facebook, and still not be assured her secrets were safe. How could she be sure that insurance paperwork, credit or bank records, or websites visited online did not somehow betray such conditions? The information could end up in the hands of a profiler like Acxiom or a scraper linking online handles to real identities.[88] As one victim complained, "I can't be the watchdog all the time."[89] Moreover, in a world where the "Internet of things," integrated databases, and video search makes privacy in public spaces obsolete, it's hard to imagine how to "opt out" of surveillance.[90] There is simply not enough time in the day to

---

[86] Cathy O'Neil, *Fighting the Information War (But Only on Behalf of Rich People)*, MATHBABE (Dec. 11, 2012), http://mathbabe.org/2012/12/11/fighting-the-information-war-but-only-on-behalf-of-rich-people/.

[87] Julie E. Cohen, *What Privacy is For*, 126 HARVARD L. REV. 1904, 1905 (2010); Frank Pasquale, *Beyond Innovation and Competition: The Need for Qualified Transparency in Internet Intermediaries*, 104 NW. U. L. REV. 105, 151 (2010) (describing privacy as an irreducibly social practice).

[88] *See, e.g.*, Angwin & Stecklow, *supra* note 27. The entire "What They Know" series at the Wall Street Journal—dozens of articles dating back to 2010—reveals, on an almost weekly basis, commercial entities (ranging from device finger-printers to data miners to scrapers) capable of analyzing data points, re-identifying data sources, and otherwise defeating once-reasonable privacy precautions.

[89] Ylan Q. Mui, *Little-known Firms Tracking Data Used in Credit Scores*, Wash. Post (July 16, 2011), http://www.washingtonpost.com/business/economy/little-known-firms-tracking-data-used-in-credit-scores/2011/05/24/gIQAXHcWII_story.html.

[90] *See Watched: A Wall Street Journal Privacy Report*, WALL ST. J., Apr. 26, 2013, ,http://www.onthemedia.org/story/289583-future-surveillance (translated in *The Future of*

scrutinize the practices of most firms—particularly those so unique and dominant that it is exceedingly unlikely that any term will be so adverse that it justifies switching to a vastly worse alternative.[91]

IV. BEYOND NOTICE AND CONSENT: USE PROHIBITIONS AS A KEY COMPONENT OF FAIR DATA PRACTICES

We need to go beyond "notice and consent" models of privacy. Policymakers need to focus on prohibiting the use of certain types of data in certain situations. In employment or basic banking services, for instance, health data should not even enter the calculus of decisionmaking. If we as a society committed to this model, individuals would feel less pressure to avoid mentioning that they had cancer on Facebook, or meticulously trying to figure out whether a condition mentioned on PatientsLikeMe might somehow lead to a higher interest rate on credit cards.

Implementing these solutions will take some time and effort. The essence of the "Privacy Threat" assumption mentioned at the beginning of this article holds that privacy protections will inevitably gum up innovation and impede economic development. However, no one should hold the current business models of data brokers and Internet giants to be sacrosanct. If you do not want to risk being secretly categorized as a "migraine sufferer" by employers, insurers, landlords, and banks for the rest of your life just for saying "I think I may have a migraine" on a Facebook status update, you should hope the law eventually forbids important decisionmakers from using health records in their decision-making.

"De-regulationists" will characterize such rules as an assault on knowledge itself and burdensome regulation.[92] But as Paul Ohm has

---

Surveillance (Apr. 2013) ,*available at* http://www.onthemedia.org/2013/apr/26/future-surveillance/transcript/.

[91] *See* Pedro G. Leon et al., *Why Johnny Can't Opt Out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising*, CARNEGIE MELLON UNIV. CYLAB (May 10, 2010), http://www.cylab.cmu.edu/files/pdfs/tech_reports/CMUCyLab11017.pdf (illustrating a sensitive consideration of the many impediments to notice and choice in a related context).

[92] I put the term in scare quotes because any decision at the governmental level made to "regulate" or "de-regulate" has a clear-cut impact on the overall level of regulation, by firms and governments, of individual action in the economy as a whole. Presuming that

argued, if the Internet can "route around" damage, it can route around the putative distortions of regulation, too.[93] Moreover, regulation has been critical to "how existing knowledge systems have evolved, and how they are encoded and enforced."[94]

In a growing variety of contexts, rights to privacy are critical to motivate the *creation* of knowledge. For example, Jerry Kang and coauthors convincingly frame scenarios where persons will not participate in self-tracking systems that could greatly improve their (and public) health if the information generated could be used against them.[95] Why should I honestly enter in all the calories and food I eat in a day in the LoseIt app if my employer or potential life insurer can demand it and negatively judge me if I betray a weakness for blueberry donuts? More seriously, what would be the public health consequences if people became too scared to use a Google search to help themselves determine if they might have bird flu?[96]

We increasingly find ourselves needing to consider knowledge *systems*, ecologies built over time and space, rather than the efficiency or fairness of single transactions in the knowledge economy. Apps may now thrive on the basis of a dyadic trade of 99 cents and vast information collection for access, but over time may begin such

---

government itself is the sole source of the coercive power of regulation is a category mistake—corporations can impose on freedom as well.

[93] Paul Ohm, *The Myth of the Superuser: Fear, Risk, and Harm Online*, 41 U.C. DAVIS L. REV. 1327, 1373 (2008).

[94] Julie E. Cohen, *Creativity and Culture in Copyright Theory*, 40 U.C. DAVIS L. REV. 1151, 1165 (2007).

[95] *See* Kang et al., *supra* note 72, at 816-17, 820.

[96] Google Flu mines search engine queries to track real-time flu trends, yielding results much faster than the CDC's traditional epidemiological surveillance methods. It is possible that such tactics could deter people from using the tool and render it useless. The efficacy of the tool has already been called into question: while Google's tool had been accurate in the past, in early 2013 Google grossly overestimated the actual flu patterns in the U.S. due to increased public interest caused by extensive media coverage. However, early detection allows for early public health intervention, which may reduce the spread of disease and potentially save lives. *See* David Wagner, *Google Flu Trends Wildly Overestimated This Year's Flu Outbreak*, THE WIRE (Feb. 13, 2013, 4:45 PM), http://www.theatlanticwire.com/technology/2013/02/google-flu-trends-wildly-overestimated-years-flu-outbreak/62113/; *but see* Paul Ohm, *The Underwhelming Benefits of Big Data*, 161 U. PA. L. REV. 339, 340 (2013) (critiquing Google Flu Trends' limited real world impact).

invasive profiling (and manipulative tactics to gain more revenue) that individuals start withdrawing from app marketplaces. Far from menacing app providers, the FTC's nudges of mobile apps toward fair data practices are a way of preserving a "grand bargain" among all participants to preserve cooperative and mutually beneficial relationships that market exchange ideally presupposes. That common set of standards and commitments is the foundation for mutually advantageous contracts between consumers, producers, and the "prosumers" whose data is increasingly valuable to innovators.[97]

V. CONCLUSION: THE EXPANDING ROLE OF HEALTH INFORMATION POLICY

Like periodization for historians, categorization of legal disputes is, for attorneys, a classic ground of contestation. Those concerned with data about disease and treatment have primarily been concerned with the *privacy, security, and confidentiality* of *medical* records. The "threat scenarios" that motivated this legal field derived from a relatively closed set of problems regarding breach, unauthorized disclosure, and obvious embarrassments.

As digitization of health data has increased, both the terms "medical" and "privacy" are inadequate to convey the magnitude of the issues raised by the volume, variety, and velocity of the information characteristic of big data methods of collection and analysis. We now more often use the acronym EHRs (electronic *health* records) rather than EMRs (electronic *medical* records) because we realize that there are many social determinants (and indicators) of health far outside of the medical system proper.[98] Health status can also increasingly be probabilistically attributed (if not definitively discovered) with reference to records from far outside the medical

---

[97] Christian Fuchs, *Class and Exploitation on the Internet*, *in* DIGITAL LABOR 211, 217 (Trebor Scholz ed., 2013) (quoting ALVIN TOFFLER, THE THIRD WAVE 267 (1980)(defining the notion of a "prosumer" as the "progressive blurring of the line that separates producer from consumer")).

[98] See, e.g., Susan D. Hall, *IOM: Put social, behavioral data in EHRs*, FierceHealthIT, at http://www.fiercehealthit.com/story/iom-urges-collection-social-behavioral-data-ehrs/2014-04-09 (Apr. 9, 2014).

system. If you're a childless man who shops for clothing online, spends a lot on cable TV, and drives a minivan, we know certain data brokers are going to assume you weigh more than the average person.[99] And we now know, thanks to innovative reporting, that recruiters for obesity drug trials will happily pay for that analysis.[100]

Thus the concept of a sectoral health privacy law like HIPAA (and its 2009 update, HITECH) fully protecting health privacy by applying a complex set of rules to a narrow range of "covered entities" in the medical sector appears increasingly quaint. Pressure comes from the other side as well: the greater the ability of data aggregation to improve care, and the more sophisticated tools of deidentification develop, the less privacy will appear as a prime motivator of health data policy than as one of many social objectives at stake in the management of increasingly comprehensive and interconnected data systems.

The new data landscape challenges conventional understandings of optimal health information flows. Both the Privacy Threat and the Aggregation Threat are being called into question by new cooperative enterprises and integrative architectures. Moreover, individualized "Control Solutions" appear less and less realistic (and even self-defeating) as big data methods allow an ever-growing array of entities to attribute "health reputations" to individuals, regardless of how well they have tried to hide their health status. Moreover, we need to reconsider whether, as a normative matter, control is our top desideratum in data policy, given how likely it is to be exercised effectively only by a small fraction of the population.

The new big data landscape presents a kaleidoscopic array of chutes and ladders. Merely focusing policymakers' energies on "Control Solutions" while ignoring the larger context of such

---

[99] Joseph Walker, *Data Mining to Recruit Sick People*, WALL ST. J., Dec. 17, 2013, *available at* http://online.wsj.com/news/articles/SB10001424052702303722104579240140554518458. The *Journal* tries to explain these big data associations by hypothesizing that large men need minivans because they cannot fit into other vehicles. But note how easily we could also rationalize the opposite conclusion: if minivan drivers were pegged as exceptionally fit, we might hypothesize that they used the large vehicle to carry around sports equipment. We should beware post hoc rationalizations of big data correlations, particularly when we are unable to review the representativeness of the data processed or the algorithms used to process it.

[100] *Id.*

competitive behavior is a recipe for irrelevant health information law and policy. It is, in Evgeny Morozov's terms, solutionism: an orientation to problems that tends to "reach[] for the answer before the questions have been fully asked."[101] Is the goal really technologically and legally enabling fine-grained data control or do we need to address something deeper in the way our society distributes rewards and opportunities based on (factors related to) health status?

Health data solutionism tends to prioritize issues that technology can address: small, algorithmically decomposable bits of wicked problems. It will continue to captivate a small army of lawyers and computer scientists (who will, not coincidentally, be poised to take advantage of "Control Solutions" far more effectively than the rest of the population). But it is hard to see how it would have prevented the problems of the Sheltons, Catherine Taylor, or the unnamed multitude stigmatized by actual or imputed health status. Policymakers must be willing to delve far deeper into actual business and employment practices with substantive, verifiable, auditable standards of nondiscrimination and fair data practices. Our present privacy paradigm myopically encourages individuals to balance strategies of self-disclosure and concealment. We should instead focus our energies on creating a society where the dangers of such decisions are radically reduced.

[101] EVGENY MOROZOV, TO SAVE EVERYTHING, CLICK HERE: THE FOLLY OF TECHNOLOGICAL SOLUTIONISM (2013) (internal marks omitted).