

Data and Power: From Individual Consent to Societal Transparency

Frank Pasquale¹

(Discussion draft for 10/29 panel on privacy and innovation.)

As the organizers of this symposium have suggested, “technological innovation is often driven by the need to protect sensitive information in areas ranging from eHealth and eCommerce to cloud computing and national security.” Some of the brightest minds in cyberlaw have focused on such innovation.² Jerry Kang’s interdisciplinary research groups have proposed “personal data vaults” to manage the emanations of sensor networks. Jonathan Zittrain’s article on “privication” proposed that the same technologies used by copyrightholders to monitor or stop dissemination of works could be adopted by patients concerned about the unauthorized spread of health information.

These technological “self-help” measures reflect privacy law’s consent paradigm. Generally speaking, data dissemination is not deemed an invasion of privacy if it is consented to. The consent paradigm requires individuals to decide whether or not, at any given time, they wish to protect their privacy.

If individuals had enough time to manage their personal data the way they manage their checkbooks and gardens, perhaps the consent paradigm would be a good foundation for addressing public concerns about privacy. If applicants could easily bargain with would-be employers over privacy, or patients with hospitals, perhaps we could rely on them to protect their interests. But the actual occurrence of such acts of self-assertion and self-protection are rare. Given the frequently abstract benefits that privacy and reputational integrity afford, they are often traded away for competitive economic advantage.³ This process further erodes societal expectations of privacy.

A collective commitment to privacy is far more valuable than a private, transactional approach that all but guarantees a race to the bottom. If such a collective commitment does not materialize, reputation systems will only deserve trust if they become as transparent as the citizens they profile. Given corporate assertion of trade secrecy (and even privacy

¹ Schering-Plough Professor in Health Care Regulation and Enforcement, Seton Hall Law School; Visiting Fellow, Princeton Center for Information Technology Policy. For those interested in my recent work on privacy and civil liberties, these two papers discuss fusion centers and internet intermediaries: *Network Accountability for the Domestic Intelligence Apparatus*, at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1680390 (coauthored with Danielle Keats Citron); *Beyond Innovation and Competition: Qualified Transparency in Internet Intermediaries*, at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1686043. I am also working on a book entitled “The Black Box Society,” which explores how corporate assertions of trade secrecy in the finance and internet fields have imperiled privacy, fairness, and economic stability.

² Surveillance studies and STS have also considered the issue. See, e.g., Katja Vries, *Identity profiling algorithms and a world of ambient intelligence*, ETHICS AND INFORMATION TECHNOLOGY, v.12 n.1, p.71-85, March 2010; MIREILLE HILDEBRANDT, ED., PROFILING THE EUROPEAN CITIZEN; *MIT Forum on Engaging Data*, at <http://senseable.mit.edu/engagingdata/downloads.html>.

³ See, e.g., Barocas and Nissenbaum, *On Notice*, at http://senseable.mit.edu/engagingdata/papers/ED_SII_On_Notice.pdf (notice and consent in online advertising are inadequate because of “(1) the confusing disconnect between the privacy policies of online publishers and the tracking and targeting third parties with whom they contract, each of whom have their own privacy policies; (2) the fickle nature of privacy policies, which may change at any time, often with short notice, and (3) the ever-increasing number of players in the ad network and exchange space, resulting in flows of user data that are opaque to users.”).

rights),⁴ reciprocal transparency will not be easy to achieve. Nevertheless, repeated breaches, fraud, and data meltdowns in the US should provoke an alliance of socially responsible businesses in health care, finance, and internet companies to lobby the US government to set minimal standards of reciprocal transparency and auditing. Consumers can only trust innovators if they can understand what is being done with that data. As we become “transparent citizens” (as Joel Reidenberg puts it), we should demand that the corporate and governmental authors of that trend reciprocate, and become more open about the data they gather.⁵

The Broken Market for Privacy

Leading scholars have modeled privacy as a purchasable commodity: as with other products, individuals have varying preferences and abilities to pay for more or less privacy. On this economic view, firms will emerge to compete to offer more or less privacy or will provide customers with various “privacy settings” that tailor online services to optimize self-disclosure. Unfortunately, each of these assumptions is problematic, especially when we reflect on the zero-sum nature of reputational capital in many settings.

Few Internet intermediaries do (or even can) compete to offer users more privacy. The vast majority of broadband users have a choice of only one or two carriers. Even in the more competitive mobile market, it is virtually unheard of for consumers to bargain for more or less privacy, or for carriers to compete on those terms. Recent revelations about ISP deals with the NSA and “Operation Vigilant” have justifiably raised concerns about user privacy, but it’s hard to see how even the most persistent consumer could get ISP legal departments to protect them from such snooping. Indeed, just asking for such a shield might be a “tip-off” that one is a person worth watching.

Instead, carriers are beginning to compete in ways that are corrosive to privacy. As Paul Ohm has documented, “[b]roadband ISPs have. . .search[ed] for new sources of revenue. . .[by] ‘trading user secrets for cash,’ which Google has proved can be a very lucrative market.” While user protests have deterred the most abusive practices, Ohm predicts that “ISPs, faced with changes in technology, extraordinary pressures to increase revenues, and murky ethical rules, will continue aggressively to expand network monitoring.” Antitrust law has been slow to recognize privacy as a dimension of product quality, and the competition that it promotes can do as much to trample privacy as to protect it.

Intermediary competition is supposed to provide users with more companies offering more options. However, competition is based primarily on immediately experienced aspects of the service, such as price and speed. The prospect of altering the terms of service for an intermediary like Facebook or Google is beyond the ambition of almost all users.

Even intermediaries with intimate knowledge of users' communications with family and friends have tended to assert almost unlimited powers over user-generated content. The social networking site Facebook attempted to legitimate this power by creating a system that

⁴ <http://www.slate.com/id/2270956/?from=rss>.

⁵ Here I am expanding some of Jack Balkin’s proposed reforms of the “national surveillance state” to the array of private sector actors upon which it increasingly relies.

allowed its users to “vote” for changes to the terms of service before they are implemented. However, University of Cambridge researchers have released a detailed report which concludes that Facebook's system is merely “democracy theatre” with little practical effect on the company's operations.

As the use and reuse of personal information becomes more deeply rooted in internet intermediary business practices, the tension between competition and privacy becomes more pronounced. For example, if a user of one social network wants to join another, she will often be reluctant to do so because of “switching costs”; she has already invested some time and effort in creating her existing profile. The chief way of reducing those costs is to require data portability, which would allow users to take their list of contacts, applications, pictures, and other items with them when they want to leave. However, such a rule (or protocol for data storage) can render the rest of the user's social graph vulnerable to unwanted exposure on the network the user migrates to. Randal Picker has described the deep tension between competition and privacy that results: restrictions on data sharing between companies encourage companies to merge.

Given these patterns of industry practice and consumer behavior, regulation will be more effective than waiting for markets to provide varied privacy options. Given the frequently abstract “benefits” that privacy and reputational integrity afford, they are often traded away for competitive economic advantage. This process further erodes the societal expectations of privacy and accuracy that underwrite respect for reputational integrity.

From Privacy to Reputation to Reciprocal Transparency

Data dissemination can raise many concerns. First, individuals may want to quarantine some information, keeping it entirely out of the hands of some entities. Second, given a large (and growing) corpus of data regarding them, they want to be judged or evaluated on the basis of an *accurate* record of their past actions and characteristics. Finally, they want to know *how* they are being judged or evaluated. How does a transition from a consent paradigm to reciprocal transparency affect each of these goals?

- 1) **Quarantine:** Quarantining data is a primary concern of the consent paradigm. However, individuals often do not realize the multiple paths data can take in order to get into critical databases. Recently contributors to the medical website PatientsLikeMe.com found that “Nielsen Co., [a] media-research firm . . . was ‘scraping,’ or copying, every single message off PatientsLikeMe's private online forums.”⁶ Had the virtual break-in not been detected, health attributes connected to usernames (which, in turn, can often be linked to real identities) could have spread into numerous databases. A reciprocal transparency paradigm would require all those harboring health data to have some certified indication of its legitimate provenance. Such certifications would be regularly audited. Data would not be allowed to persist without certification of its provenance.
- 2) **Accuracy:** In the consent paradigm, consumers bargain with data providers for the right to future inspection of files created about them. In the reciprocal transparency paradigm, individuals have the right to inspect the files that are

⁶ <http://online.wsj.com/article/SB10001424052748703358504575544381288117888.html>

created about them, and to demand their correction.

- 3) **Reputation and Evaluation:** The consent paradigm concerns itself solely with the transfer of data. Once it has been lawfully acquired by an entity, that entity is free to do with it whatever it wants. In a regime of reciprocal transparency, the creators of reputation and evaluation systems are obliged to disclose how those systems work. While trade secrets may be kept from the public at large, secrecy is no excuse for preventing auditors, regulators, and other certified entities from examining the integrity of a reputation and evaluation system.

I expect point 3, on reputation and evaluation, to be the most controversial, so I'll make that case in a bit more detail. Consider the following scenarios involving clandestine computing power:

- A) **Black Box Health Simulations and Medical Decision Support:** "[A] computer model . . . called Archimedes . . . [is] a kind of SimHealth: a vast compendium of medical knowledge drawn from epidemiological data, clinical trials, and physician interviews . . . laboriously translated into differential equations over the past decade. Those equations . . . [may] successfully reproduce the complex workings of human biology — down to the individual chambers of a simulated person's virtual heart."⁷

[But the system's inventor's] "secretive habits are . . . troubling, according to [the] director of the Diabetes Center at Massachusetts General Hospital. "[He] has 10,000 variables and differential equations describing everything from blood sugar to office furniture. . . . But it's never quite clear what they are or how they interact. All the calculations happen inside a black box. . . . [T]here's no way to tell whether the model's underlying assumptions are right."

- B) **Credit Scoring & Reputation Creation:** Credit scoring puts consumers in a related dilemma. You may know your score, but it's unclear how it is calculated. A bad score may cost a borrower tens or hundreds of thousands of dollars over the life of a loan. More importantly, auto and home insurers and even hospitals are looking at credit scores. A company called Recorded Future, backed by the CIA and Google, promises to predict which employees are most likely to quit. All these entities zealously protect their trade secrets.
- C) **The Spread of Corporate Data to Governmental Entities:** Warrantless wiretapping is just the tip of an iceberg of new domestic intelligence programs that rely on private companies to act as "big brother's little helpers."⁸ As the "national surveillance state" has grown, entities called "fusion centers" have collapsed the traditional distinctions between law enforcement and foreign wars, and between government surveillance and corporate data practices. Inadequate oversight over fusion centers has led to significant civil liberties abuses. Yet their practices and

⁷ http://www.wired.com/magazine/2009/11/ff_archimedes/all/1

⁸ Washington Post Series, TopSecretAmerica.com.

priorities remain hidden not only from citizens, but also from traditional agents of accountability. Given law enforcement exceptions to medical privacy laws and regulations, it should come as little surprise that the government claims that “a 2005 law authorizes it to monitor and record all prescription drug use by all citizens via so-called “Prescription Drug Monitoring Programs.”⁹

Institutionalizing Reciprocal Transparency

To avoid misuse of personal data, and to catch faulty processing of data, regulators need to move beyond the consent paradigm of privacy to a much more robust system of accountability. The law of information offers many tools for deflecting current practices of corporate secrecy that prevent individuals from understanding the “digital persons” they appear to be in online dossiers.

In some cases, law simply destroys secrecy, requiring disclosure. This is a strategy of “total transparency.” Other times, law only requires the keeper of a secret to disclose it to a judge or a limited number of individuals who are themselves sworn to secrecy. I call this practice “qualified transparency.” It has been a time-honored part of the law of trade secrets, where a “protective orders” allow judges and parties’ attorneys to examine trade secrets during litigation. The Dodd-Frank Act of 2010 establishes “qualified transparency” in finance, giving regulators at the newly created OFR (called a “CIA for Finance”) important ways to assess systemic risk while keeping proprietary trading books closed to the public.

When a black box is impregnable (or imperfectly transparent), law can attempt a “quarantine” strategy, limiting the reach of the information it produces. Most timidly, law may merely require annotation, warning individuals that the credit score, search ranking, or derivative they are using a) merely reflects a subjective opinion, or b) is too complex to be understood by anyone other than the entity that created it.

Credit score regulation may be leading the way here. In popular books like Ian Ayres’s *Super Crunchers* and Stephen Baker’s *The Numerati*, data-driven decisionmaking is celebrated as a cornerstone of future advances in productivity. However, the individual who is an *object* of such “supercrunching” may fear that a crucial decision about her is being made on the basis of a misunderstanding—an unfair reduction of a complex person to a number. These concerns have led some lawmakers to prevent certain private sector decisionmakers from even looking at credit scores.

Though a credit score is computed via proprietary algorithms protected as trade secrets, it is widely treated as a fair and objective evaluation of an individual's creditworthiness. The industry remains opaque, with scored individuals unable to determine the exact consequences of late payments, changes in location, or other decisions. Several disturbing reports have alleged racial and other inappropriate influences on credit scores. Because of concerns about their unreliability and unfairness, use of credit scores has been regulated by forty-eight states, and five states prohibit automobile insurers from even considering them. Minnesota lawmakers passed a bill prohibiting the use of “medical credit scores” at hospitals, but it was vetoed by the governor.

⁹ Glenn Greenwald, at <http://www.cato-unbound.org/2010/08/09/glenn-greenwald/the-digital-surveillance-state-vast-secret-and-dangerous/>.

Unaccountable data usage in the employment sphere should also attract attention. One survey found that, in the U.S., “as many as 50% of employers and 77% of job recruiters concerned about alcohol/drug abuse, violence, and similar problems check out potential employees on the Web.” Such digital background checks would be much more difficult in Finland, which has forbidden employers from using Google results (among other unauthorized information sources) in evaluating potential applicants.

While that type of quarantine approach is unlikely to be widely adopted, employment law could make the *use* of black boxes less mysterious. Requiring important decisionmakers to reveal the online sources they use in order to evaluate applicants would be an important first step. More ambitiously, the US should expand the Fair Credit Reporting Act into a Fair Reputation Reporting Act. Such a law would require lenders, insurers, employers, and educators to reveal the particular information they found out about an applicant to that applicant after any decision is made. These disclosures would be a first step toward educating individuals about the digital personas that internet intermediaries construct about them online.

Finally, journals and other arbiters of knowledge should apply ideas of reciprocal transparency in the academic world. As the Roundtable on Data and Code Sharing at Yale Law School declared this year, reproducible research should be a cornerstone of the new computational science.¹⁰ The roundtable stated that “Reproducible computational research, in which all details of computations--code and data--are made conveniently available to others, is a necessary step in addressing the current credibility crisis in computational science.” Auditability and replicability are also key to assuring that reputation and evaluation systems outside the academy respect individuals’ concerns about accurate and fair data processing.¹¹

Inexorable technological and social forces will continue to put more data out of the control of the individuals whom the data (putatively) describes. Even if this process ends many traditional conceptions of privacy, it can be harnessed to generate the accountability and fair outcomes that privacy law was designed to promote, if “watchers” reciprocate and allow themselves to be better watched.

¹⁰ <http://www.computer.org/portal/web/csdl/abs/html/mags/cs/2010/05/mcs2010050008.htm>. (I was a part of this discussion.)

¹¹ Respected voices in the domestic intelligence community have also made this point about systems like the ISE. MARKLE TASK FORCE ON INFORMATION SECURITY IN THE INFORMATION AGE, IMPLEMENTING A TRUSTED INFORMATION SHARING ENVIRONMENT USING IMMUTABLE AUDIT LOGS TO INCREASE SECURITY, TRUST, AND ACCOUNTABILITY 1 (2006) (arguing that, if immutable audit logs of fusion centers are audited regularly, misconduct might be discovered, wrongdoers might be held responsible, and similar misuses might be deterred); SHANE HARRIS, THE WATCHERS 190 (2009) (explaining that in the proposed Total Information Awareness (TIA) system, John Poindexter “proposed an ‘immutable audit trail,’ a master record of every analyst who had used the TIA system, what data they touched, and what they’d done with it . . . to spot suspicious patterns of use Poindexter wanted to use TIA to watch the watchers.”).