

RELATIONAL BIG DATA

Karen E.C. Levy*

“Big Data” has attracted considerable public attention of late, garnering press coverage both optimistic and dystopian in tone. Some of the stories we tell about big data treat it as a computational panacea—a key to unlock the mysteries of the human genome, to crunch away the problems of urban living, or to elucidate hidden patterns underlying our friendships and cultural preferences.¹ Others describe big data as an invasive apparatus through which governments keep close tabs on citizens, while corporations compile detailed dossiers about what we purchase and consume.² Like so many technological advances before it, our stories about big data generate it as a two-headed creature, the source of both tremendous promise and disquieting surveillance. In reality, like any complicated social phenomenon, big data is both of these, a set of heterogeneous resources and practices deployed in multiple ways toward diverse ends.³

* Ph.D. Candidate, Princeton University.

1. See Michael Specter, *Germes Are Us*, NEW YORKER, Oct. 22, 2012, at 32, available at http://www.newyorker.com/reporting/2012/10/22/121022fa_fact_specter (human genome); Alan Feuer, *The Mayor’s Geek Squad*, N.Y. TIMES, Mar. 24, 2013, at MB1, available at <http://www.nytimes.com/2013/03/24/nyregion/mayor-bloombergs-geek-squad.html> (city services); Nick Bilton, *Looking at Facebook’s Friend and Relationship Status Through Big Data*, N.Y. TIMES BITS BLOG (Apr. 25, 2013), <http://bits.blogs.nytimes.com/2013/04/25/looking-at-facebooks-friend-and-relationship-status-through-big-data> (interpersonal relationships).

2. Two much-discussed recent examples are the National Security Agency’s wide-ranging PRISM data collection program, Charlie Savage et al., *U.S. Confirms Gathering of Web Data Overseas*, N.Y. TIMES, June 7, 2013, at A1, available at <http://www.nytimes.com/2013/06/07/us/nsa-verizon-calls.html>, and the revelation that Target collected purchasing data that predicted the pregnancy of a teenage girl before her family knew about it, Charles Duhigg, *Psst, You in Aisle 5*, N.Y. TIMES, Feb. 19, 2012, at MM30, available at <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>.

3. The slipperiness of the definition here isn’t helped by the vagueness around whether big data is *data* or *practice*—the millions or billions of pieces of information being examined or the methodological tools for its examination. Much of the “new” information to which big data refers isn’t actually new (we have always had a genome); what *is* new is our capacity to collect and analyze it.

I want to complicate matters further by suggesting another way in which data has become big: data now *mediate our day-to-day social relationships* to an unprecedented degree. This other big data revolution relies on the proliferation of new data collection and analysis tools that allow individuals to track easily, quantify, and communicate information about our own behaviors and those of others. This type of big data arguably touches more of us more directly than the big data practices more commonly discussed, as it comes to reshape our relationships across multiple domains of daily life.

In this sense, data is big not because of the number of points that comprise a particular dataset, nor the statistical methods used to analyze them, nor the computational power on which such analysis relies. Instead, data is big because of the depth to which it has come to pervade our personal connections to one another. A key characteristic of this flavor of big data, which I term “relational”⁴ (more on this in a moment) is *who* is doing the collection and analysis. In most big data stories, both dreamy and dystopian, collection and analysis are *top-down*, driven by corporations, governments, or academic institutions. In contrast, relational big data is collected and analyzed by *individuals*, inhabiting social roles (as parents, friends, etc.) *as a means for negotiating social life*. In other words, we can understand big data not simply as a methodological watershed, but as a fundamental social shift in how people manage relationships and make choices, with complex implications for privacy, trust, and dynamics of interpersonal control.

Another notable distinction is the multiplicity of sources of relational big data. While most analyses of social big data focus on a few behemoth forums for online information-seeking and interaction, what Zeynep Tufekci describes as “large-scale aggregate databases of imprints of online and social media activity”⁵—Google, Facebook, Twitter, and the like—I suggest that “datafication” extends well beyond these digital presences, extending into diverse domains and relying on multiple dispersed tools, some of which are household names and some of which never will be.

In the rest of this Essay, I flesh out the idea of relational big data by describing its conceptual predecessor in economic sociology. I suggest a few domains in which data mediate social relationships and how interactions might

4. My use of the term “relational” here is distinct from the computational meaning of the word (i.e., relating to the structure of a database). I also do not mean “relational” in the sense of merely having to do with social networks and communications, though other big data analysis is based on such associations. See, e.g., Katherine J. Strandburg, *Freedom of Association in a Networked World: First Amendment Regulation of Relational Surveillance*, 49 B.C.L. REV. 741 (2008).

5. Zeynep Tufekci, *Big Data: Pitfalls, Methods, and Concepts for an Emergent Field*, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2229952 (Mar. 7, 2013). A great deal of scholarly work has investigated how digital communication forums like Facebook and Twitter mediate interactions both on- and off-line. See, e.g., Alice Marwick & danah boyd, *I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience*, 13 NEW MEDIA AND SOC’Y 114 (2010).

change around it. I then consider what analytical purchase this flavor of big data gets us regarding questions of policy in the age of ubiquitous computing.

I. WHAT'S RELATIONAL ABOUT DATA?

To say that big data is *relational* borrows a page from economic sociology, particularly from the work of Viviana Zelizer.⁶ As its name implies, economic sociology broadly examines the social aspects of economic life, from how markets are structured to the development of money. One of Zelizer's seminal contributions to the field is the idea that economic exchanges do "relational work" for people: through transactions, people create and manage their interpersonal ties. For example, individuals vary the features of transactions (in search of what Zelizer calls "viable matches" among interpersonal ties, transactions, and media) in order to differentiate social relationships and create boundaries that establish what a relationship *is* and *is not*. (Consider, for instance, why you might feel more comfortable giving a coworker a gift certificate as a birthday present rather than cash.) Thus, to construe transactions merely as trades of fungible goods and services misses a good part of what's interesting and important about them.

I suggest that we should do for data practices what Zelizer does for economic practices: we should consider that people use data to create and define relationships with one another. Saying that data practices are relational does more than simply observe that they occur against a background of social networks; rather, people *constitute and enact* their relations with one another *through* the use and exchange of data.⁷ Consider, for example, a person who monitors the real-time location of her friends via a smartphone app designed for this purpose. By monitoring some friends but not others, she differentiates among her relationships, defining some as closer. By agreeing to share their locations, her friends communicate that they have no expectation of privacy (to her) as to where they are, perhaps suggesting that they trust her. The acts of sharing and monitoring say a lot about the nature of the relationship; focusing only on the locational data itself, as much big data analysis does, ignores the social negotiations taking place via data practices.

Big data is, at heart, a social phenomenon—but many of the stories we tell about it reduce people to mere data points to be acted upon. A relational framework is appealing because it puts people, their behaviors, and their relationships at the center of the analysis as active agents. Big data and its attendant

6. Viviana Zelizer, *How I Became a Relational Economic Sociologist and What Does That Mean?*, 40 POL. & SOC'Y 145 (2012) [hereinafter Zelizer, *Relational Economic Sociologist*]; Viviana Zelizer, *Pasts and Futures of Economic Sociology*, 50 AM. BEHAV. SCIENTIST 1056 (2007).

7. Zelizer similarly contrasts her relational perspective with the previous "embeddedness" approach in economic sociology. Zelizer, *Relational Economic Sociologist*, *supra* note 6, at 162.

practices aren't monoliths; they are diverse and socially contingent, a fact which any policy analysis of big data phenomena must consider.

II. BIG DATA DOMAINS

Data pervade all kinds of social contexts, and the tools available to gather and use data vary tremendously across them. In what types of relationships do data circulate? I touch on a few here.

Children and families. Technologies for data gathering and surveillance within families are proliferating rapidly. A number of these involve monitoring the whereabouts of family members (often, though not always, children). One such product, LockDown GPS, transmits data about a vehicle's speed and location so parents can easily monitor a teen's driving habits. The system can prevent a car from being restarted after it's been shut off, and parents are immediately notified of rule violations by e-mail. The system purports to "[put] the parent in the driver's seat 24 hours a day, from anywhere in the world."⁸

A number of other products and apps (like FlexiSpy, Mamabear, My Mobile Watchdog, and others) allow individuals to monitor data like the calls a family member receives, the content of texts and photos, real-time location, Facebook activity, and the like, with or without the monitored party being aware of it. And not all intra-family monitoring is child-directed: a number of products market themselves as tools for tracking down untrustworthy spouses,⁹ while others detect such behaviors as whether an elder parent has taken his or her medicine.¹⁰

Communities and friendships. Jeffrey Lane's ethnographic account of three years spent living with Harlem youth describes how they manage diverse relationships with friends, rivals, and authority figures using social media.¹¹ An abundance of other tools enable us to relate to our communities through data by, for instance, finding friends in physical space (Find My Friends), selecting local businesses to patronize (Yelp), or "checking in" to physical locations (Foursquare).

The workplace. The use of productivity metrics to manage employees is far from new, but the proliferation of tools for doing so introduces data into new

8. *Family Protection*, LOCKDOWN SYSTEM, INC., <http://www.lockdownsystems.com/basic-page/family-protection> (last visited Aug. 29, 2013).

9. See, e.g., Sophie Curtis, 'Boyfriend Tracker' App Banished from Google Play, TELEGRAPH (Aug. 22, 2013), <http://www.telegraph.co.uk/technology/news/10259516/Boyfriend-Tracker-app-banished-from-Google-Play.html>.

10. See, e.g., *How the Philips Medication Dispensing Service Works*, PHILIPS, http://www.managemypills.com/content/How_PMD_Works (last visited Aug. 29, 2013).

11. Jeffrey Lane, Presentation on Code-Switching on the Digital Street at the American Sociological Association Annual Meeting (August 12, 2013); see also danah boyd & Alice Marwick, *Social Steganography: Privacy in Networked Publics* (May 9, 2011) (unpublished manuscript), available at <http://www.danah.org/papers/2011/Steganography-ICAVersion.pdf>.

kinds of employment relationships. Parents can monitor a caretaker's behavior via nanny cam. Fast-growing workplace wellness monitoring programs frequently use health indicators and behavioral data (derived, for instance, from a digital pedometer) to let employers and insurers keep tabs on the health of their workforce.¹² Highly mobile employees like truck drivers, who traditionally are accorded a good deal of occupational autonomy, are increasingly monitored via fleet management and dispatch systems that transmit data about their driving habits, fuel usage, and location to a central hub in real time—practices that have engendered deep concerns about driver privacy and harassment.¹³

Self-monitoring. Finally, individuals increasingly use electronic data gathering systems to control their *own* behavior. The Quantified Self “movement” is the most acute example of this—Quantified Selfers monitor their own biophysical, behavioral, and environmental markers in efforts to measure progress toward health and other goals.¹⁴ Even among those who would not identify with such a movement, a number of self-tracking systems have recently emerged on the consumer electronics market (for example, the FitBit and Nike FuelBand), while popular services like 23AndMe, Mint, and Daytum facilitate tracking of genetic information, personal finance, and myriad other types of data. Even when monitoring is self-directed, however, these data can impact interpersonal relationships (for example, by facilitating comparison and competition within one's personal networks).¹⁵

In many areas of life, then, individuals use data gathering and analysis tools to manage their relationships with one another in a variety of ways, only a few of which I mention here. In some cases, data help people to *control* the actions of others by serving as a digital site of accountability for action, poten-

12. Workplace health monitoring practices are not without critics; CVS recently faced criticism from privacy advocates for its announcement that workers would be fined \$600 per year if they failed to disclose health metrics to the company's insurer. See Christine McConville, *CVS Presses Workers for Medical Information*, BOS. HERALD (Mar. 19, 2013), http://bostonherald.com/business/healthcare/2013/03/cvs_presses_workers_for_medical_information.

13. For instance, new proposed regulations that would require truckers' work hours to be electronically monitored have been challenged due to the possibility that motor carriers will use the technology to harass drivers. See *Owner-Operators Indep. Drivers Ass'n v. Fed. Motor Carrier Safety Admin.*, 656 F.3d 580 (7th Cir. 2011).

14. See Gary Wolf, *The Data-Driven Life*, N.Y. TIMES, May 2, 2010, at MM38, available at <http://www.nytimes.com/2010/05/02/magazine/02self-measurement-t.html>. Of course, this phenomenon is not entirely new; self-monitoring has long been an element of many wellness efforts before the digital age (e.g., analog diet and exercise tracking). But, digital tools markedly increase the scale and depth of monitoring programs, as well as facilitating the use of such data for interpersonal competition.

15. A recent anecdote in the *New Yorker* describes husband-and-wife FitBit users; the husband checks the wife's activity stats at the end of the day and will suddenly take the dog for a walk, while the wife, knowing what her husband is up to, paces around the house while he's gone to prevent him from “winning.” Susan Orlean, *The Walking Alive*, NEW YORKER, May 20, 2013, at 44, 47, available at http://www.newyorker.com/reporting/2013/05/20/130520fa_fact_orlean.

tially diminishing the need for social trust (for instance, monitoring a teen's car may effectively undermine the need for parent-child trust by creating a seemingly objective record of compliance or noncompliance with parental rules). In others, technologies facilitate *competition* in relationships: employment metrics are commonly publicized to encourage intra-workforce competition, and many health-centric data services allow and encourage users to compete with peers and strangers. Such competition is not merely an externality of the use of these devices, but a central reason why these techniques can be effective. Third, data practices may help individuals to *distinguish* between relationships and send desired signals to one another (e.g., as suggested earlier, adding certain friends but not others to a find-my-friends service). The meanings and effects of data practices vary considerably within and across life domains.

III. POLICY, PRIVACY, IMPLICATIONS

Big data poses big problems for privacy,¹⁶ which are only compounded by the relational framework I suggest. Top-down data collection programs create the need for strong civil liberties protections, due process, and assurances of data integrity. But the privacy interests implicated by relational big data are bound up in particular social contexts,¹⁷ no single piece of legislation or court ruling would prove a useful tool to protect them.

Instead, it is likely that some privacy interests implicated by relational big data may figure into existing legal frameworks governing personal relationships (for instance, workplace harassment, or tort claims like invasion of privacy) or in some cases via domain-specific rules, such as laws governing the use of medical or genetic information.¹⁸ Gathered data may also come to legal use as evidence, substantiating an alibi or providing proof of a fact like vehicle speed. But in most cases, interpersonal privacy intrusions facilitated by relational data-gathering tools fall outside the realm of legal redress, precisely because the law is traditionally hesitant to get involved in the minutiae of personal relationships.

Despite the fact that law doesn't provide a clear approach, policymakers and privacy scholars still have much to gain from thinking about relational data practices. The ubiquity of interpersonal data-gathering activities helps us understand people as *both subjects and objects* of big data regimes, not just data points. When people collect and use data to constitute their relationships

16. See Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63 (2012).

17. See HELEN NISSENBAUM, *PRIVACY IN CONTEXT: TECHNOLOGY, POLICY, AND THE INTEGRITY OF SOCIAL LIFE* (2010).

18. For instance, laws like Health Insurance Portability and Accountability Act (HIPAA) of 1996, Pub. L. No. 104-91, 110 Stat. 1936, and Genetic Information Nondiscrimination Act (GINA) of 2008, Pub. L. No. 110-233, 122 Stat. 881, protect privacy interests in health-related and genetic information.

with one another, social norms around accountability, privacy, veracity, and trust are likely to evolve in complex ways.

In addition, thinking about individuals this way may be instructive when considering public responses to top-down surveillance. For instance, although recent revelations about the NSA's PRISM surveillance program (in which essentially every major technology provider secretly supplied consumer communications to the NSA) excited much outrage among academics and civil libertarians, news of the program's existence engendered a comparatively tepid response from the general public.¹⁹ Part of the reason may be that we have become docile²⁰ in light of the ubiquity and pervasiveness of data gathering across domains of daily life. Relational data practices may instill in the public a tolerance for watching and being watched, measuring and being measured, that leads us to abide additional surveillance without much complaint.

19. Poll data suggest that sixty-six percent of Americans support the government's collection of Internet data via the PRISM program. Brett LoGiurato, *The NSA's PRISM Program is Shockingly Uncontroversial with the American Public*, BUS. INSIDER (June 17, 2013), <http://www.businessinsider.com/prism-surveillance-poll-nsa-obama-approval-2013-6>.

20. Michel Foucault famously described how disciplinary techniques create "docile bodies" accustomed to further discipline. For instance, he observed that schools operated as "pedagogical machine[s]," analogous to institutions like the factory and the prison: by inculcating disciplinary systems in students, schools prepare young subjects to encounter similar techniques in other realms. MICHEL FOUCAULT, *DISCIPLINE AND PUNISH: THE BIRTH OF THE PRISON* 172 (1977).