HHS Public Access

Author manuscript

Am Sociol Rev. Author manuscript; available in PMC 2024 February 06.

Published in final edited form as:

Am Sociol Rev. 2017 October ; 82(5): 977–1008. doi:10.1177/0003122417725865.

Big Data Surveillance: The Case of Policing

Sarah Brayne[a]

[a]The University of Texas at Austin

Abstract

This article examines the intersection of two structural developments: the growth of surveillance and the rise of "big data." Drawing on observations and interviews conducted within the Los Angeles Police Department, I offer an empirical account of how the adoption of big data analytics does—and does not—transform police surveillance practices. I argue that the adoption of big data analytics facilitates amplifications of prior surveillance practices and fundamental transformations in surveillance activities. First, discretionary assessments of risk are supplemented and quantified using risk scores. Second, data are used for predictive, rather than reactive or explanatory, purposes. Third, the proliferation of automatic alert systems makes it possible to systematically surveil an unprecedentedly large number of people. Fourth, the threshold for inclusion in law enforcement databases is lower, now including individuals who have not had direct police contact. Fifth, previously separate data systems are merged, facilitating the spread of surveillance into a wide range of institutions. Based on these findings, I develop a theoretical model of big data surveillance that can be applied to institutional domains beyond the criminal justice system. Finally, I highlight the social consequences of big data surveillance for law and social inequality.

Keywords

police; big data; inequality; crime; law

In the past decade, two major structural developments intersected: the proliferation of surveillance in everyday life and the rise of "big data." Emblematic of the expansion of surveillance (Lyon 2003; Marx 2016; Rule 2007) is the rapid growth of the criminal justice system since 1972 (Carson 2015; Garland 2001; Wakefield and Uggen 2010; Western 2006). At the same time, facilitated by the mass digitization of information, there has been a rise in the computational analysis of massive and diverse datasets, known as "big data." Big data analytics have been taken up in a wide range of fields, including finance, health, social science, sports, marketing, security, and criminal justice. The use of big data in police surveillance activities is the subject of contentious debate in policy, media, legal, regulatory, and academic circles. However, discourse on the topic is largely speculative, focusing on the *possibilities*, good and bad, of new forms of data-based surveillance. The technological capacities for surveillance far outpace empirical research on the new

Corresponding Author: Sarah Brayne, Assistant Professor, Department of Sociology, The University of Texas at Austin, 305 E. 23rd St., A1700, Austin, TX 78712-1086, sbrayne@utexas.edu.

data landscape. Consequently, we actually know very little about how big data is used in surveillance activities and to what consequence.

This article provides a case study of a large urban police department—the Los Angeles Police Department (LAPD)—to investigate the relationship between big data analytics and surveillance. In particular, it asks whether and how the adoption of big data analytics transforms police surveillance practices. Moreover, it investigates implications of new surveillance practices not only for policing, but also for law, social inequality, and research on big data surveillance in other institutions. On the one hand, big data analytics may be a rationalizing force, with potential to reduce bias, increase efficiency, and improve prediction accuracy. On the other hand, use of predictive analytics has the potential to technologically reify bias and deepen existing patterns of inequality.

To shed light on the social dimensions of surveillance in the age of big data, I draw on original interview and observational data collected during fieldwork over the course of two and a half years with the LAPD. As an agency at the forefront of data analytics, the Department serves as a strategic site for understanding the interplay between technology, law, and social relations. I provide one of the first on-the-ground accounts of how the growing suite of big data systems and predictive analytics are used for surveillance within an organization, unpacking how in some cases, the adoption of big data analytics is associated with mere *amplifications* in prior practices, but that in others, it is associated with fundamental *transformations* in surveillance activities. I argue there are five key ways in which the adoption of big data analytics is associated with shifts in practice to varying degrees: (1) discretionary assessments of risk are supplemented and quantified using risk scores; (2) data are increasingly used for predictive, rather than reactive or explanatory, purposes; (3) the proliferation of automated alerts makes it possible to systematically surveil an unprecedentedly large number of people; (4) datasets now include information on individuals who have not had any direct police contact; and (5) previously separate data systems are merged into relational systems and include data originally collected in other, non-criminal-justice institutions. These shifts made possible by big data have implications for inequality, law, and organizational practice in a range of institutional domains.

THE INTENSIFICATION OF SURVEILLANCE

Surveillance is ubiquitous in modern societies (Giddens 1990; Lyon 1994, 2003, 2006, 2015; Marx 1974, 2002, 2016; Rule 1974). The practice of surveillance involves the collection, recording, and classification of information about people, processes, and institutions (Foucault 1977; Haggerty and Ericson 2000; Lyon 2003; Marx 2016). A wide range of scholars highlight the growing pervasiveness of surveillance, referring to the emergence of “mass surveillance” (Rule 1974) and “surveillance societies” (Lyon 1994). Although it has received more attention in recent decades, surveillance as a practice is not new. Processes of surveillance can be traced back to at least the sixteenth century, during the appearance of the nation-state (Marx 2016), through the transatlantic slave trade (Browne 2015), bureaucratization, rationalization, and modern management in the nineteenth and twentieth centuries (Braverman 1974; Rule 1974; Weber 1978), and as an axiomatic accompaniment to risk management practices in the twentieth century (Ericson and Haggerty 1997). The

attacks on September 11th, 2001, further stimulated and legitimated the expansion of surveillance. Widely viewed as a case of information sharing failure in the intelligence community, 9/11 encouraged an alignment of actors with vested interests in enhancing surveillance operations (Ball and Webster 2007; Lyon 2015), spurred an infusion of tax dollars for development of new surveillance sensors and data mining programs to produce strategic intelligence (Gandy 2002), and accelerated the convergence of previously separate surveillance systems into a "surveillant assemblage" (Deleuze and Guattari 1987; Haggerty and Ericson 2000).

Surveillance scholars have documented a quantitative increase in surveillance, arguing that surveillance is one of the major institutional dimensions of modern societies (Ball and Webster 2007; Giddens 1990; Lyon 1994, 2003; Marx 1988, 2016). Although surveillance is growing in all areas of society, its penetration is unevenly distributed (Fiske 1998). Some individuals, groups, areas, and institutions are surveilled more than others, and different populations are surveilled for different purposes (Lyon 2003). On the one hand, there is a deepening of surveillance of "at-risk" groups, such as parolees and individuals on public assistance (Gilliom 2001; Gustafson 2011; Soss, Fording, and Schram 2011), who can increasingly be tracked across institutional boundaries. On the other hand, emerging "dragnet" surveillance practices—meaning those that collect data on everyone, rather than merely individuals under suspicion—result in increased monitoring of groups "previously exempt from routine surveillance" (Haggerty and Ericson 2000:606; see also Angwin 2014; Lyon 2015). Surveillance is therefore now both *wider* and *deeper*: it includes a broader swath of people and can follow any single individual across a greater range of institutional settings.

Surveillance is increasingly technologically mediated, and emergent technologies make it possible at an unprecedented scale (Ericson and Haggerty 1997; Lyon 1994; Marx 2016). With the development of computing, mass surveillance emerged alongside mass communication (Rule 1974). The mass digitization of information enables much of what Marx terms the "new surveillance": the "scrutiny of individuals, groups, and contexts using technical means to extract or create information" (Marx 2016:20). Although the goals of traditional and new surveillance are similar, the means are different. Whereas traditional surveillance is inductive, involving the "close observation, especially of a suspected person" (Oxford American Dictionary of Current English 1999), and relying on the unaided senses, new surveillance is more likely to be applied categorically, deductive, remote, low visibility or invisible, involuntary, automated, preemptive, and embedded into routine activity (Marx 2002, 2016). In fact, new forms of systematic surveillance have become quotidian to the point that it is now an unavoidable feature of everyday life (Ball and Webster 2007; Lyon 2003; Marx 2016). Consider the extent to which surveillance is a requisite of participating in today's world (Ball and Webster 2007): to use a bank, send an e-mail, obtain medical care, make a phone call, travel on a highway, or conduct an Internet search, individuals leave digital traces that are recorded and saved (see also Foucault 1977; Haggerty and Ericson 2000; Lyon 2003).

Technologically mediated surveillance has become routine organizational practice in a wide range of public and private domains. It is a tool of general governance (Lyon 2003),

a basic prerogative of individuals in all sorts of private and public institutions, and a tool to accomplish "goals that meet particular interests" (Ericson and Haggerty 2006:22). Surveillance scholars have accordingly extended the core concerns of surveillance from policing and military contexts to other institutions, including finance, commerce, labor, health, education, insurance, immigration, and activism (Lyon 2003; Marx 2016; Rule 2007). According to David Lyon (2015:68–69), surveillance in each of these institutions, "cannot be understood without a sense of how the quest for 'big data' approaches are becoming increasingly central."

RISE OF BIG DATA

Big data is an emerging modality of surveillance. A wide range of organizations—from finance to healthcare to law enforcement—have adopted big data analytics as a means to increase efficiency, improve prediction, and reduce bias (Christin 2016). Despite its take-up, big data remains an ambiguous term whose precise definition can vary across fields and institutional contexts. Drawing on previous definitions (e.g., Laney 2001; Lazer and Radford 2017; Mayer-Schönberger and Cukier 2013), the working definition of big data used in this research is that it is a data *environment* characterized by four features: it is vast, fast, disparate, and digital. First, big data analytics involve the analysis of large amounts of information, often measured in petabytes and involving tens of millions of observations. Second, big data typically involves high frequency observations and fast data processing. Third, big data is disparate—it comes from a wide range of institutional sensors and involves the merging of previously separate data sources. Fourth, big data is digital. The mass digitization of records facilitates the merging and sharing of records across institutions, makes storage and processing easier, and makes data more efficient to analyze and search remotely. These four characteristics are not limited to any one institutional context, and they enable the use of advanced analytics—such as predictive algorithms or network analysis—and complex data display—such as topical-, temporal-, or geo-analysis. For purposes of sociological research on the topic, this definition shifts the focus from features of the data itself to the social processes that give rise to big data collection and analysis (i.e., the data environment). For example, instead of focusing on the "variety" of big data (one of the "3 Vs" in Laney's [2001] definition), the focus here is on the disparate institutional data sources big data is gathered from.

There are numerous theories for why such a wide range of institutions adopted big data surveillance as an organizational practice, most of which fit into one of two theoretical perspectives: the technical/rational perspective and the institutional perspective. Both perspectives are premised on the notion that organizations are self-interested (Scott 1987), but actors within organizations may adopt big data analytics in response to different pressures. According to the technical perspective, big data is a means by which organizational actors improve efficiency through improving prediction, filling analytic gaps, and more effectively allocating scarce resources. By contrast, the institutional perspective (DiMaggio and Powell 1983; Meyer and Rowan 1977) questions the assumption that organizational structures stem from rational processes or technical imperatives (Scott 2004). Instead, it highlights the role of culture, suggesting organizations operate in technically ambiguous fields in which they adopt big data analytics not because of empirical evidence

that it *actually* improves efficiency, but in response to wider beliefs of what organizations *should* be doing with big data (Willis, Mastrofski, and Weisburd 2007; see also Kling 1991 on computerization). In other words, using big data may confer legitimacy. If other institutions are marshalling big data and algorithmic predictions for decision-making—rather than relying on human discretion—there may be institutional pressure to conform.

Big data makes possible new forms of classification and prediction using machine learning algorithms. Applications of big data analytics range from spam and fraud detection to credit scoring, insurance pricing, employment decisions, and predictive policing. Although much of the appeal of algorithms lies in their replacement of human decision-making with automated decisions, this study highlights the ways in which humans remain integral to the analytic process.

In the big data environment, individuals contribute to a growing trove of data as they go about their daily lives (Ball and Webster 2007; Garland 2001). Every time people make a purchase using a credit card, drive through a tollbooth, or click on an advertisement online, they leave a digital trace (Rule 2007). The adoption of digital information and communications technologies transformed previously paper files and face-to-face data collection (Marx 1998), making it possible for records "initially introduced with limited intentions" to be "developed, refined and expanded to deal with new problems and situations" (Innes 2001:8). Function creep—the tendency of data initially collected for one purpose to be used for another often unintended or unanticipated purpose (Innes 2001)—is a fundamental component of the big data surveillant landscape.

BIG DATA SURVEILLANCE: THE CASE OF POLICING

This article provides a case study of policing, one of the many organizational contexts in which the use of big data surveillance has grown. More generally, criminal justice surveillance has increased dramatically in the United States in the past four decades. It has expanded at all levels, including incarceration (Travis, Western, and Redburn 2014), parole and probation (Bonczar and Herberman 2014), and policing (Carson 2015). For example, the Violent Crime Control and Law Enforcement Act of 1994 provided funds to hire 100,000 new police officers, and the Homeland Security Act of 2002 committed over 17 billion dollars for state and local governments to fund local law enforcement agencies (Roush 2012). More recently, federal funds have been targeted at improving and expanding law enforcement's use of technology. For example, the Smart Policing Initiative—a consortium of the Bureau of Justice Assistance, local police departments, and researchers—provides federal funds to more than 30 local law enforcement agencies (including the LAPD) to support new data-driven practices.

The use of data for decision-making in criminal justice is not new. In 1928, Ernest Burgess of the Chicago School designed an actuarial model that predicted the probability of parolees' reoffending (Harcourt 2006). In the courts, quantification was embedded into legal practices in the 1970s and 1980s through sentencing guidelines (Espeland and Vannebo 2007). In the past three decades, the criminal justice system experienced a shift toward "actuarial justice" (Feeley and Simon 1992), in which actors use criteria derived from risk management (Lyon

2003) to estimate probabilities of criminal risk (Ericson and Haggerty 1997). That said, although actuarial methods have existed in corrections and the courts for almost a century (Feeley and Simon 1992; Harcourt 2006; Lyon 2003), data-driven decision-making has become systematically incorporated into law enforcement practices only in recent decades.

In the 1970s, the dominant police patrol model was reactive (Reiss 1971), involving random patrols, rapid responses to 911 calls, and reactive investigations (Sherman 2013). However, practitioners and researchers became increasingly aware that these strategies had little effect on crime, catalyzing a shift from reactive to more proactive, evidence-based forms of policing, such as hot spots policing (Braga and Weisburd 2010; Sherman, Gartin, and Buerger 1989). In 1994, CompStat—a management model linking crime and enforcement statistics—was established in New York City (Weisburd et al. 2003). CompStat quickly spread to other cities, including Los Angeles in 2002, as a managerial model for identifying crime patterns, quantifying and incentivizing police activity, and directing police resources. The attacks on 9/11 spurred the development of "intelligence-led policing" (Ratcliffe 2008). Viewing local law enforcement agencies as actors on the front lines of the domestic war against terror (Waxman 2009), federal agencies provided considerable funding to local law enforcement agencies to collect, analyze, share, and deploy a wide range of new data. In 2008, William Bratton, then-Chief of the LAPD (and former Commissioner of the New York City Police Department) began working with federal agencies to assess the viability of a more predictive approach to policing. Today, predictive analytics are used for a wide range of law enforcement–related activities, including algorithms predicting when and where future crimes are most likely to occur (Perry et al. 2013), network models predicting individuals most likely to be involved in gun violence (Papachristos, Hureau, and Braga 2013), and risk models identifying law enforcement officers most likely to engage in at-risk behavior (U.S. Department of Justice 2001 [2015]).

What explains the proliferation of big-data-driven decision-making in organizations generally, and law enforcement specifically? Much like in other institutional domains, it has the potential to improve both efficiency and accountability. It may improve the prediction and preemption of behaviors by helping law enforcement deploy resources more efficiently, ultimately helping prevent and intercept crimes, thus reducing crime rates. Data-driven policing also holds potential as an accountability mechanism and response to criticisms organizations are facing over discriminatory practices. For example, in response to police violence, nationwide movements such as Black Lives Matter have brought racial tensions to the forefront of demands for police reform. Data-driven policing is being offered as a partial antidote to racially discriminatory practices in police departments across the country (e.g., see White House Police Data Initiative 2015).

However, although part of the appeal of big data lies in its promise of less discretionary and more objective decision-making (see Porter's [1995] work on mechanical objectivity; see also Espeland and Vannebo 2007; Hacking 1990), new analytic platforms and techniques are deployed in preexisting organizational contexts (Barley 1986, 1996; Kling 1991) and embody the purposes of their creators (boyd and Crawford 2012; Gitelman 2013; Kitchin 2014). Therefore, it remains an open empirical question to what extent the adoption of advanced analytics will reduce organizational inefficiencies and inequalities, or serve

to entrench power dynamics within organizations. The present study sheds light on these questions and helps us understand the changing relationship between quantification, prediction, and inequality.

LIMITATIONS TO EXISTING LITERATURE

There are five key limitations to the existing literature related to big data surveillance. First, most work on actuarialism was written before big data analytics took hold. Second, although there is strong theoretical work in surveillance studies, how big data surveillance plays out on the ground remains largely an open empirical question. Third, the majority of sociological research on criminal justice surveillance focuses on the experiences and outcomes of individuals under surveillance, rather than the surveilling agents themselves. Therefore, offering an organizational perspective may generate new insight about this mediating level of analysis. Fifth, although there is a strong body of work demonstrating that marking someone in the criminal justice system is consequential for life outcomes and patterns of inequality (Becker 1963; Brayne 2014; Kohler-Hausmann 2013; Pager 2007; Rios 2011), we know relatively little about whether and how the marking process has changed in the age of big data. Consequently, there is a dearth of theoretically informed empirical research on the relationship between surveillance, big data, and the social consequences of the intersection of the two forces.

Many of these gaps in the policing context can be attributed to practical constraints—it is difficult for researchers to secure the degree of access to police departments necessary to obtain in-depth qualitative data on day-to-day police practices. Although classic police ethnographies exist (e.g., Bittner 1967; Manning and Van Maanen 1978; Wilson 1968), there have been only a handful of in-depth studies within police departments since data analytics became an integral part of police operations (for early exceptions, see Ericson and Haggerty 1997; Manning 2011; Moskos 2008; Skogan 2006; Willis et al. 2007). Although these studies offer important insight into the use of CompStat and crime mapping, they predate algorithmic policing. Consequently, we still know little about how big data policing is exercised in practice.

This article has three aims. First, it draws on unique, original data to analyze how a law enforcement organization conducts big data surveillance. Second, it forwards an original theoretical framework for understanding the changes—and continuities—in surveillance practices associated with the adoption of big data analytics that can be applied to other institutional domains. Finally, it highlights implications of big data surveillance for law and social inequality.

FIELDWORK

Over the course of two and a half years, I conducted a qualitative case study of the Los Angeles Police Department (LAPD). The LAPD is the third-largest local law enforcement agency in the United States, employing 9,947 sworn officers and 2,947 civilian staff (Los Angeles Police Department 2017). The Department covers an area of almost 500 square miles and a population of almost four million people. It consists of four bureaus—Central,

South, Valley, and West—which are divided into a total of 21 geographic areas. There are also two specialized bureaus, Detective and Special Operations.

I conducted interviews and observations with 75 individuals, and conducted between one and five follow-up interviews with a subsample of 31 individuals to follow how certain technologies were disseminated and information was shared throughout the Department. Interviewees included sworn officers of various ranks and civilian employees working in patrol, investigation, and crime analysis. I was able to gain analytic leverage by exploiting different adoption temporalities within and between divisions during my fieldwork. Not all divisions adopted big data surveillance at the same time. Rather, there was considerable variation in whether and when different big data technologies were adopted in the area and specialized divisions.[1] For example, I was able to conduct interviews and observations in divisions that were not using predictive policing and other big data surveillant technologies at the beginning of my fieldwork, but were by the end. This variation enabled me to observe actual changes in practice, but also to talk to respondents in patrol, investigation, and analysis roles about how they interpreted their work changing in light of big data analytics. I also interviewed individuals in specialized divisions—including Robbery-Homicide, Information Technology, Records and Identification, Fugitive Warrants, Juvenile, Risk Management, and Air Support—and at the Real-Time Crime Analysis Center (see Figure 1).

Additionally, I conducted observations on ride-alongs in patrol cars and a helicopter to study how officers deploy data in the field. I also shadowed analysts as they worked with data, observing them responding to queries from detectives and supervisors and proactively analyzing data for patrol, investigations, and crime analysis.

To supplement my research within the LAPD, I interviewed individuals within the L.A. County Sheriff's Department (LASD), as it is an integral part of the broader ecosystem of public services in the region. In addition, I conducted interviews at the Joint Regional Intelligence Center (JRIC), the "fusion center" in Southern California. Fusion centers are multiagency, multidisciplinary surveillance organizations by state or local agencies that received considerable federal funding from the Department of Homeland Security and the Department of Justice (Monahan and Palmer 2009). JRIC is one of 78 federally funded fusion centers established across the country in the wake of 9/11. Individuals at JRIC conduct data collection, aggregation, and surveillance in conjunction with other fusion centers and agencies, including, but not limited to, the Department of Homeland Security (DHS), the Federal Bureau of Investigations (FBI), the Central Intelligence Agency (CIA), and Immigration and Customs Enforcement (ICE). I also conducted observations at surveillance industry conferences and interviewed individuals working at technology companies that design analytic platforms used by the LAPD, including Palantir and PredPol, and individuals working in federal agencies in Washington, DC, to understand how data on criminal and noncriminal activity are shared across agencies. I supplemented my fieldwork with archival research of law enforcement and military training manuals and surveillance

[1]For example, one division in the valley began using PredPol (predictive policing) in 2012, and nine other divisions followed suit between then and March 2015.

industry literature. Triangulating across various sources of data provided the analytic leverage necessary to better understand how law enforcement uses big data in theory, how they use it in practice, and how they interpret and make meaning out of its changing role in daily operations.

Site Selection

I selected the LAPD as a strategic site for studying big data surveillance because it is an agency at the forefront of data analytics. The LAPD invests heavily in its data collection, analysis, and deployment capacities, and offers international training sessions on how law enforcement can better harness big data. Therefore, practices within the Department may forecast broader trends that may shape other law enforcement agencies in the coming years.

In addition to being one of the largest law enforcement agencies in North America, there are additional, contextual reasons why the LAPD is on the leading edge of data analytics. The first factor relates to external pressures for transparency and accountability. The LAPD was involved in a number of high-profile scandals in the 1990s, including the Rampart Scandal[2] and the now infamous Rodney King beating, which led to investigations exposing an expansive web of corruption, training deficiencies, and civil rights violations within the Department. In response, the Department of Justice entered into a consent decree[3] with the LAPD from 2001 to 2009 that mandated, among other things, the creation and oversight of a new data-driven employee risk management system, TEAMS II. The legacy of the decree extends beyond employee risk management; it led to more information sharing and data-driven decision-making within the organization in general.

The second factor is the influence of state legislative decisions concerning offender management. In the wake of *Brown v. Plata*[4] and the associated order to dramatically reduce the prison population, the California Legislature passed AB 109, a bill that shifted the responsibility of supervising released non-violent, non-serious, non-sex offenders from state to local law enforcement and county probation officers. It also outsourced compliance checks to local law enforcement agencies, including the LAPD and LASD. As a result, local law enforcement agencies were responsible for approximately 500 additional individuals released into L.A. County each month. Therefore, they needed a means by which to efficiently stratify the post-release community supervision population according to risk, necessitating risk modeling and interagency data integration efforts across the region.

A third relevant factor to the LAPD's use of big data is the availability and adoption of new data integration technologies. In 2011, the LAPD began using a platform designed by Palantir Technologies. Palantir was founded in 2004 and has quickly grown into one of the premier platforms for compiling and analyzing massive and disparate data by law enforcement and intelligence agencies. Originally intended for use in national defense,

2. More than 25 officers in Rampart Division's special operations anti-gang unit, C.R.A.S.H., were investigated or charged, and over 100 criminal cases were overturned due to police misconduct.
3. A consent decree is a binding court order memorializing an agreement between parties in exchange for an end to a civil litigation or a withdrawal of a criminal charge.
4. Brown v. Plata is a 2011 U.S. Supreme Court decision holding that the overcrowding of California prisons and lack of access to adequate healthcare violated prisoners' Eighth Amendment constitutional rights.

Palantir was initially partially funded by In-Q-Tel, the CIA's venture capital firm. Palantir now has government and commercial customers, including the CIA, FBI, ICE, LAPD, NYPD, NSA, DHS, and J.P. Morgan. JRIC (the Southern California fusion center) started using Palantir in 2009, with the LAPD following shortly after. The use of Palantir has expanded rapidly through the Department, with regular training sessions and more divisions signing on each year. It has also spread throughout the greater L.A. region: in 2014, Palantir won the Request for Proposals to implement the statewide AB 109 administration program, which involves data integration and monitoring of the post-release community supervision population.

CHANGES ASSOCIATED WITH ADOPTION OF BIG DATA ANALYTICS

To what extent does the adoption of big data analytics change police surveillance? Based on my fieldwork, I argue that in some cases, the adoption of big data analytics is associated with mere *amplifications* in prior surveillance practices, but in others, it is associated with fundamental *transformations* in surveillance activities and daily operations. I empirically demonstrate five key ways in which the adoption of big data analytics is associated with shifts in surveillance practices to varying degrees. First, law enforcement supplements officers' discretionary assessments of risk with quantified risk scores. Second, there is an increase in the use of data analytics for predictive—rather than reactive or explanatory—purposes. Third, there is a proliferation in alert-based systems, which facilitates the passive, systematic surveillance of a larger number of individuals than is possible with traditional query-based systems. Fourth, the threshold for inclusion in law enforcement databases is lower, now including individuals who have not had direct police contact. Finally, previously separate data systems are merged into relational systems, making it possible for the police to use data originally collected in other, non–criminal justice contexts.

I offer an original conceptual framework for understanding the continuities and changes associated with the adoption of big data analytics within the police organization. Figure 2 depicts this framework and illustrates the migration of traditional police practices toward big data surveillance. Each line represents a continuum of surveillance practices, from traditional to big data surveillance. The five shifts in practice do not represent discrete either/or categories, but rather are better understood as continuous gradations of varying degrees between the extreme values of traditional and big data surveillance. The length of the black lines represents the degree of transformation in surveillance practices associated with the use of big data. For example, the first two shifts—from discretionary to quantified risk assessment, and explanatory to predictive analytics—are not particularly transformative; rather, they represent quantified recapitulations of traditional surveillance practices. By contrast, the last two shifts—the inclusion of data on individuals with no direct police contact, and from institutions typically not associated with crime control—represent fundamental transformations in surveillance activities. The shift from query-based systems to automated alerts is a moderate shift, representing, in part, an elaboration of existing practices, and in part, a new surveillance strategy. I will analyze each of these shifts in the following sections.

The shift from traditional to big data surveillance is associated with a migration of law enforcement operations toward intelligence activities. The basic distinction between law enforcement and intelligence is as follows: law enforcement typically becomes involved once a criminal incident has occurred. Legally, the police cannot undertake a search and gather personal information until there is probable cause. Intelligence, by contrast, is fundamentally predictive. Intelligence activities involve gathering data; identifying suspicious patterns, locations, activity, and individuals; and preemptively intervening based on the intelligence acquired. Before discussing the findings, one caveat is worth noting: the migration of law enforcement toward intelligence was in its nascency before the use of big data, in part due to Supreme Court decisions dismantling certain criminal protections. Technically, the Fourth Amendment makes unreasonable searches and seizures illegal in the absence of probable cause. However, in practice, decisions such as *Terry v. Ohio* and *Whren v. United States* made it easier for law enforcement to circumvent the barrier of probable cause, ultimately contributing to the proliferation of pretext stops. In other words, the Supreme Court's dismantling of probable cause catalyzed the migration of law enforcement toward intelligence, and the adoption of big data analytics facilitated and accelerated this shift.

The Quantification of Individual Risk

The first shift in police practice is the quantification of civilians according to risk. Quantified knowledge is supplementing officers' experiential knowledge through the implementation of a new point system: Operation LASER (Los Angeles' Strategic Extraction and Restoration program). The program began in 2011 and was funded through the Smart Policing Initiative, a national initiative encouraging local police departments and researchers to use evidence-based, data-driven tactics. The strategy includes place-based and offender-based models. The offender-based strategy was implemented in a low-income, historically high-crime division in South Bureau. It is premised on the idea that a small percentage of high-impact players are disproportionately responsible for most violent crime. Therefore, identifying and focusing police resources on the "hottest" individuals should be an efficient means of reducing crime.[5]

The strategy begins by plotting crimes in the division. The Crime Intelligence Detail (CID), which is composed of three sworn officers and a civilian crime analyst, identifies a problem crime, which in this division is often armed robbery. Next, the CID shifts their unit of analysis from crimes to individuals. They gather intelligence daily from patrols, the Parole Compliance Unit, field interview (FI) cards (police contact cards), traffic citations, release from custody forms, crime and arrest reports, and criminal histories to generate a list of "chronic offenders," who are each assigned a point value and given a numerical rank according to that value. Individuals are assigned five points for a violent criminal history, five points for known gang affiliation, five points for prior arrests with a handgun, and five points if they are on parole or probation. One officer explained:

[5]To date, one study has evaluated the efficacy of Operation LASER (Uchida and Swatt 2015). It found a reduction in crime in reporting districts that adopted both person-based and location-based approaches. The program has not yet been subject to external evaluation; the authors are the President of and Senior Research Associate at Justice and Security Strategies, who designed Operation LASER.

> We said ok, we need to decide who's the worst of the worst . . . we need something to pull them apart. So this was the important one, and this is really what gives the importance of FI-ing someone [filling out a field interview card] on a daily basis instead of just saying, okay, I saw that guy hanging out, I'm gonna give him two weeks and I'll go FI him again. *It's one point for every police contact*.[6]

As illustrated in Figure 3, FI cards include personal information such as name, address, physical characteristics, vehicle information, gang affiliations, and criminal history. On the back of the card, there is space for officers to include information on persons with the subject and additional intelligence.

FIs are key intelligence tools for law enforcement and were one of the first data sources integrated into Palantir. When entered into the system, every FI is tagged with the time, date, and geo-coordinates. Officers are trained to pull out an FI card and "get a shake" as soon as they interact with someone in the field. One supervisor described how he uses it "to tag all the personal information I can get . . . these things come into play later on in ways you could never even imagine." Similarly, a software engineer explained how little pieces of data that might seem unsuspicious at the time of collection can eventually be pulled together to create useful intelligence: "It's a law enforcement system where that citation can, the sum of all information can build out what is needed." In addition to inputting the *content* of the cards, a captain explained there is an incentive to simply "get them in the system" as entities that future data points can be linked to.

Because point values are largely based on police contact, an important question emerges: What are grounds for police contact? Merely being identified as a chronic offender does not constitute reasonable suspicion or probable cause. However, used in conjunction with Palantir, FIs represent a proliferation of data from police–civilian interactions that law enforcement does not need a warrant to collect. When I asked an officer to provide examples of why he stops people with high point values, he replied:

> Yesterday this individual might have got stopped because he jaywalked. Today he mighta got stopped because he didn't use his turn signal or whatever the case might be. So that's two points . . . you could conduct an investigation or if something seems out of place you have your consensual stops.[7] So a pedestrian stop, this individual's walking, "Hey, can I talk to you for a moment?" "Yeah what's up?" You know, and then you just start filling out your card as he answers questions or whatever. And what it was telling us is who is out on the street, you know, who's out there not necessarily maybe committing a crime but who's active on the streets. You put the activity of . . . being in a street with maybe their violent background and one and one might create the next crime that's gonna occur.

The point system is path dependent; it generates a feedback loop by which FIs are both causes and consequences of high point values. An individual having a high point value is

[6] Block quotes are drawn from audiotaped, transcribed interviews.

[7] Consensual stops may be conducted at any time when the police lack the "specific and articulable facts" (*Terry v. Ohio* 392 U.S. at 21) that justify detention or arrest.

predictive of future police contact, and that police contact further increases the individual's point value.

The CID also creates work-ups, referred to as "Chronic Violent Crime Offender Bulletins." Individuals on these bulletins are not necessarily "wanted" nor do they have outstanding warrants for their arrest. Rather, it is an "information only" bulletin that includes physical descriptors and oddities, gang affiliation, criminal history, parole/probation status, vehicles, frequented areas, and law enforcement contacts. The goal of these bulletins is to give officers what they refer to as "situational awareness." Officers previously had to rely exclusively on their direct knowledge of a case and specific criminal networks, but by creating a list and disseminating bulletins, the point system and associated bulletins broadens previously particularized police familiarity of individuals on the street.

Ideally, one officer explained, they could put one officer on every individual on the list and "odds are [you're] probably going to find them committing another crime." However, the police operate in an organizational context with resource constraints; respondents frequently referenced budget cuts and personnel shortages. Therefore, instead of having one officer on every chronic offender, officers engage in what I term "stratified surveillance": differentially surveilling individuals according to their risk score. An officer explained:

> [We] utilize undercover operations, or undercover units and . . . then sit our surveillance on some of the higher point offenders and just watch them on a daily basis. . . . And you start building either, you know, there's two ways of looking at it. Either kind of conducting your investigation to see if maybe there was a crime that had just been committed. Or, "We know who you are, you know, I just called you Johnny, I've never really met you before, but I know who you are now," so maybe it's put in his mind, "Oh, they're on to me, they know who I am."

This excerpt sheds light on the multiple purposes of stratified surveillance, including ongoing intelligence gathering and deterrence through signaling to individuals on the street that they are being tracked by law enforcement.

Why did the police turn to the point system in this division? In the words of one officer,

> The code of federal regulations. They say you shouldn't create a—you can't target individuals especially for any race or I forget how you say that. But then we didn't want to make it look like we're creating a gang depository of just gang affiliates or gang associates. . . . We were just trying to cover and make sure everything is right on the front end.

Other respondents echoed this sentiment, explaining the strategy was adopted, in part, as a legal compliance mechanism.

The point system is a form of quantified policing, but it is not dramatically different from its discretionary predecessor. As indicated by the low degree of transformation in Figure 2, it is largely a quantified recapitulation of traditional surveillance practices.

Shift from Reactive to Predictive Analytics

Historically, policing was mostly reactive. Patrol officers used to spend much of their time "chasing the radio." In the early 1980s, faced with evidence that reactive strategies were ineffective at reducing crime, there was a paradigm shift toward more proactive, problem-oriented policing strategies, including hot spots policing. Predictive policing is an extension of hot spots policing, made possible by the temporal density of big data (i.e., high-frequency observations). In 2012, the LAPD began using software designed by PredPol, a predictive policing company. PredPol uses a proprietary algorithm[8] predicated on the near-repeat model, which suggests once a crime occurs in a location, the immediate surrounding area is at increased risk for subsequent crime. PredPol uses three types of inputs—past type, place, and time of crime—to identify areas where future crime is most likely to occur. Predictive policing is expanding rapidly within the Department; as of March 2015, it had disseminated to 10 divisions.[9]

Officers receive printouts at the beginning of their shift that show 500 by 500 square-foot boxes overlaying small areas of division maps. Patrol officers are encouraged to spend time in predictive boxes, a strategy referred to as "risk-based deployment." Deployment is based on available time, such as when officers are not responding to calls or "booking a body." Officers record their self-reported minutes in the predictive boxes on their in-car computers. Although "data drives deployment," what the police do once in the predictive box, and how long they stay there, remains within their discretion.

One supervisor explained that by relying on data, rather than human interpretation of crime patterns, it helps him deploy his resources more efficiently:[10]

> There's an emotional element to it, and you think right now with crime being this low, a cluster could be three or four crimes. Clusters used to be 10, 12 crimes. Now three or four and they jump on it, you know. So, there could be overreaction. Because, there's, you know, I mean it's a human doing it. And they cannot sort out what's noise.

Officers were quick to emphasize the continued importance of their own expertise. When discussing predictive policing, most patrol officers said some version of the following statement made by a sergeant on a ride-along: "I already know where the crime's at." Part of this sentiment may stem from officers' concern that the use of algorithms represents a form of deskilling, devaluing their local and experiential knowledge and threatening their professional autonomy. In that vein, one captain described a typical exchange with his officers:

---

8. PredPol's algorithm was published by Mohler and colleagues in 2015.

9. In line with other law enforcement agencies, the LAPD is a hierarchical organization and employees are usually subject to tight managerial control. However, there was more between-division variation in the use of algorithms than I originally expected. Big data technologies were not adopted in the 21 area divisions and the specialized divisions at the same time, largely due to the autonomy granted to captains in each division during the early stages of algorithmic policing. Entrepreneurial captains who were early adopters used algorithmic techniques largely of their own volition, but as predictive policing was piloted in more divisions, one captain explained to me that he was starting to feel pressure to use it in his division, because he did not want to be the last to sign on.

10. In a randomized controlled field trial, Mohler and colleagues (2015) found PredPol's algorithm outperforms crime analysts predicting crime, and that police patrols using algorithmic forecasting led to significant reductions in crime volume. The algorithm has not yet been subject to external evaluation; the authors include co-founders and stockholders of PredPol.

> They're like, "You know what, I know where the crime's occurring." . . . And I show them the forecast and they say, "Okay, so [at intersection], I know there are crimes, I could have told you that. I've been working here 10 years! There's always crime there." I go, "Okay, you're working here 10 years on that car, why is there still crime there if you're so knowledgeable?"

Despite some within-department conflict over their efficacy, PredPol outputs still informed where some officers drove during their uncommitted time. For example, when driving back from booking an individual at the station, a sergeant I was with decided to drive to an area not known to him for being high-crime, because he thought the location of the PredPol box was odd and he wanted to see what was going on there. In other words, predictive policing outputs sometimes—but not always—acted as a substitute for localized experiential knowledge.

A related but distinct reason why officers contest predictive policing is because they believe it places officers themselves under greater surveillance. For example, when we arrived at a crime scene on my first ride-along, I was surprised to see an officer manually type our location on his laptop. Considering how technologically advanced the Department was in other ways, I assumed cars' locations would be tracked automatically. When I asked the officer why he manually placed himself at the scene, he explained that although every police unit was equipped with an automatic vehicle locator (AVL) that pings the vehicle's location every five seconds, they were not turned on because of resistance from LAPD union representatives.[11]

Shift from Query-Based to Alert-Based Systems

The shift from query-based to alert-based systems represents, in part, an extension of existing practices and, in part, a fundamental transformation in surveillance activities. By "query-based systems," I mean databases to which users submit requests for information in the form of a search. A familiar example of a query is when a police officer runs a license plate during a traffic stop. In alert-based systems, by contrast, users receive real-time notifications (alerts) when certain variables or configurations of variables are present in the data. The shift from query-based to alert-based systems—which is made possible by high frequency data collection—has implications for the relational structure of surveillance.

Consider the following example: all warrants in L.A. County can be translated into object representations spatially, temporally, and topically in Palantir. Through tagging, users can add every known association that warrant has to people, vehicles, addresses, phone numbers, documents, incidents, citations, calls for service, ALPR readings, FIs, and the like. Officers and analysts can then set up alerts by putting a geo-fence around an area and requesting an alert every time a new warrant is issued within the area. Warrants are but one example; users can request alerts for any data points related to the entity they are interested in (e.g., calls for service, involvement in or witnesses to a traffic accident, ALPR [automatic license plate reader] readings, FIs, and so on). Using a mechanism in Palantir similar to an RSS feed, officers can be automatically notified of warrants or events involving specific individuals (or

11. After protracted negotiations, AVLs were turned on in Central Bureau in March 2015.

matching descriptions of individuals), addresses, or cars directly on their cell phone. Prior to automated alerts, law enforcement would know individuals' real-time location only if they were conducting 1:1 surveillance, received a tip, or encountered them in person.

Real-time notifications can be useful in operational planning. An interviewee who worked at the fusion center described how if he is about to conduct a search of a house, he can draw a fence around the house and receive notifications about risks such as whether a known gang associate lived in the home, if there was a gun registered in the house next door, or if there was a warrant for assault with a deadly weapon issued down the street.

Alerts can also be used to break down information silos within the Department. LAPD's jurisdiction is almost 500 square miles. Therefore, individual detectives may not able to connect crime series that occur across different divisions. One captain explained:

> Let's say I have something going on with the medical marijuana clinics where they're getting robbed. Okay? And it happens all over, right? But I'm a detective here in [division], I can put in an alert to Palantir that says anything that has to do with medical marijuana plus robbery plus male, black, six foot.

He continued, "I like throwing the net out there, you know? Throw it out there, let it work on it while you're doing your other stuff, you know?" Relatedly, an interviewee in Robbery-Homicide Division described a pilot project in which automated data grazing can flag potential crime series that span jurisdictional boundaries and are therefore difficult for any one person to identify. He said, "You could get an alert that would say, you know what, your case is pretty similar to this case over in Miami." If the case reaches a "merit score" (i.e., a threshold at which a certain configuration of variables is present), the system flags the cases as similar. The system matches on fields such as suspect description, license plate, type of weapon, cause of death, motive, type of crime, and M.O., such as "what kind of bindings were used . . . or was there torture involved? What type of trauma has occurred? Was there, you know, was there some type of symbolic activity?" Although the matching process is automated, decisions about what parameters the system matches on remain within the discretion of individuals at ViCAP, a unit of the FBI.

That said, the use of alerts represents not just a scaling up of existing police practices, but also a fundamental transformation in how patrol officers and investigators generate case knowledge. Under the traditional surveillance model, alerts about hot incidents and suspects are sent out from dispatch centers. However, by exploiting variation in divisions that did and did not use place- and person-based predictive policing—and divisions that started using big data during the course of my fieldwork—I was able to observe the automation of alerts and relative lack of human intermediation in broadcasting out these alerts or conducting data grazing.

It is worth noting that alert-based systems are supplementing, rather than replacing, query-based systems. Searches are still critical features of law enforcement information systems. In fact, one of the transformative features of big data systems is that *queries themselves are becoming data*. One detective explained:

> I queried the system a certain way and then another person queried the system a certain way . . . we were looking for something very similar in our query, and so even though the data may not have connected the two, the queries were similar. Yeah, so then it will be able to say hey, listen, there's an analyst in San Francisco PD that ran a very similar query as to yours and so you guys might be looking for the same thing.

A different detective explained how when he searches someone's name in one national system, he can see the number of times that name has been queried by other people. When I asked why he would want to know how many times someone's name has been queried, he replied that "if you aren't doing anything wrong," the cops are not going to be looking you up very many times over the course of your life. He continued: "Just because you haven't been arrested doesn't mean you haven't been caught." In other words, in auditable big data systems, queries can serve as quantified proxies for suspiciousness.

Lower Database Inclusion Thresholds

The last two shifts in practice represent the most fundamental transformations in surveillance activities. Law enforcement databases have long included information on individuals who have been arrested or convicted of crimes. More recently, they also include information on people who have been stopped, as evidenced by the proliferation of stop-and-frisk databases. However, as new data sensors and analytic platforms are incorporated into law enforcement operations, the police increasingly utilize data on individuals who have not had any police contact at all. Quotidian activities are being codified by law enforcement organizations. One way this is occurring is through network analysis. Figure 4 is a de-identified mockup I asked an employee at Palantir to create based on a real network diagram I obtained from an officer in the LAPD. The person of interest, "Guy Cross," is an individual with a high point value. An LAPD officer explained, "with the [Palantir] system . . . I click on him and then [a] web would spread out and show me the phones that he's associated with and the cars."

Radiating out from "Guy Cross," who has direct police contact, are all the entities he is related to, including people, cars, addresses, and phone numbers. Each line indicates how they are connected, such as by being a sibling, lover, cohabiter, co-worker, co-arrestee, or listed on a vehicle registration. The network diagram illustrates only one degree of separation, but networks can expand outward to as many degrees of separation as users have information and can tie in with other networks. To be in what I call the "secondary surveillance network," individuals do not need to have direct law enforcement contact; they simply need to have a link to the central person of interest. Once individual relationships are inputted and social networks are built into the system, individuals can be "autotracked," meaning officers can receive real-time alerts if individuals in the network come into contact with the police or other government agencies again.

The Automatic License Plate Reader (ALPR) is another example of a low-threshold "trigger mechanism" (Tracy and Morgan 2000) that results in more widespread inclusion in a database. ALPRs are dragnet surveillance tools; they take readings on everyone, not merely those under suspicion. Cameras mounted on police cars and static ALPRs at intersections

take two photos of every car that passes through their line of vision—one of the license plate and one of the car—and records the time, date, and GPS coordinates (see Figure 5). Law enforcement–collected ALPR data can be supplemented with privately collected ALPRs, such as those used by repossession agents. ALPR data give the police a map of the distribution of vehicles throughout the city and, in some cases, may enable law enforcement to see an individual's typical travel patterns. For example, an analyst used ALPR data to see that a person of interest was frequently parked near a particular intersection at night, explaining to me that this intersection is likely near that person's residence or "honeycomb" (hideout).

There are several ways to use ALPR data. One is to compare them against "heat lists" of outstanding warrants or stolen cars. Another strategy is to place a geo-fence around a location of interest in order to track cars near the location. For example, after a series of copper wire thefts in the city, the police found the car involved by drawing a radius in Palantir around the three places the wire was stolen from, setting up time bounds around the time they knew the thefts occurred at each site, and querying the system for any license plates captured by ALPRs in all three locations during those time periods.

However, the most common use of ALPRs is simply to store data for potential use during a future investigation. For example, one sergeant described a "body dump" (the disposal of a dead body) that occurred in a remote location near a tourist attraction where there was an ALPR. By searching ALPR readings within the time frame that police determined the body was disposed, they captured three plates—one from Utah, one from New Mexico, and one from Compton. The sergeant explained that assuming the Compton car was most likely to be involved, they ran the plate, saw the name it was registered under, searched the name in CalGang (gang database), saw that the individual was affiliated with a gang currently at war with the victim's gang, and used that information to establish probable cause to obtain a search warrant, go to the address, find the car, search the car for trace evidence, and arrest the suspect.

Although LAPD and Palantir employees frequently told me that to be "in the system," a person needed to have had criminal justice contact, the use of network diagrams and the inclusion of ALPR data in the Palantir platform offer clear examples in which individuals with no criminal justice contact are included in law enforcement databases.

Institutional Data Systems Are Integrated

Finally, the proliferation of digitized records makes it possible to merge data from previously separate institutional sources into an integrated, structural system in which disparate data points are displayed and searchable in relation to one another, and individuals can be cross-referenced across databases. This integration facilitates one of the most transformative features of the big data landscape: the creep of criminal justice surveillance into other, non–criminal justice institutions. Function creep—the phenomenon of data originally collected for one purpose being used for another—contributes to a substantial increase in the data police have access to. Indeed, law enforcement is following an institutional data imperative (Fourcade and Healy 2017), securing routine access to a wide range of data on everyday activities from non-police databases. Before Palantir, officers and

analysts conducted predominantly one-off searches in “siloed” systems: one to look up a rap sheet, another to search a license plate, another to search for traffic citations, and so on. The Palantir platform integrates disparate data sources and makes it possible to quickly search across databases.

Expressing his faith in the Department’s investment in the platform, one captain told me, “We’ve dumped hundreds of thousands into that [Palantir]. . . . They’re gonna take over the world. . . . I promise you they’re gonna take over the world.” During my fieldwork, there were more than 1,300 trained Palantir users in the region. New data sources are incorporated regularly, including information collected by the Department, external data collected by other government agencies, and privately collected data the Department purchases. Remarking on the growth, one captain said:

> I’m so happy with how big Palantir got. . . . I mean it’s just every time I see the entry screen [Figure 6] where you log on there’s another icon about another database that’s been added . . . they now have been working with Palantir to develop a database of all the foreclosure properties . . . they just went out and found some public data on foreclosures, dragged it in, and now they’re mapping it where it would be relative to our crime data and stuff.

The Palantir platform allows users to organize and visualize structured and unstructured data content (e.g., e-mails, PDFs, and photos) through “tagging,” the process of labeling and linking objects and entities to identify emerging relationships. By tagging objects and entities—including, but not limited to, persons, phone numbers, addresses, documents such as law enforcement reports or tips and leads, and calls for service—and displaying the data spatially, temporally, or topically, users can see data points in context and make new connections.

Another important interagency data integration effort is the initiative to create an Enterprise Master Person Index (EMPI) in L.A. County. L.A. EMPI would create a single view of a client across all government systems and agencies; all of an individual’s interactions with law enforcement, social services, health services, mental health services, and child and family services would be merged onto one unique ID. Although interviewees working in the county’s information technology office stated the explicit motivation behind the initiative was to improve service delivery, such initiatives effectively serve the latent function of extending the governance and social control capacities of the criminal justice system into other institutions.

I encountered several other examples of law enforcement using external data originally collected for non–criminal justice purposes, including data from repossession and collections agencies; social media, foreclosure, and electronic toll pass data; and address and usage information from utility bills. Respondents also indicated they were working on integrating hospital, pay parking lot, and university camera feeds; rebate data such as address information from contact lens rebates; and call data from pizza chains, including names, addresses, and phone numbers from Papa Johns and Pizza Hut. In some instances, it is simply easier for law enforcement to purchase privately collected data than to rely on in-house data because there are fewer constitutional protections, reporting requirements,

and appellate checks on private sector surveillance and data collection (Pasquale 2014). Moreover, respondents explained, privately collected data is sometimes more up-to-date.

It is worth noting that such data integration is not seamless. Merging data from different sources and creating interoperable systems is part of the invisible labor that makes big data analytics possible. One civilian employee lamented, "You always forget about the data guy . . . [the] guy that does all the dirty work is usually forgotten. I'm that guy." Moreover, efforts at acquiring new data were not received evenly throughout the Department. A vocal minority of interviewees explained they did not believe leadership was fully thinking through the implications of collecting such a wide range of new data. A civilian employee complained about the seduction of new technology, saying: "We tend to just say, 'Let's just go for the sexy tool,' right? . . . We just never think about to what end." He added,

> Maybe we shouldn't collect this information. Maybe we shouldn't add consumer information. Maybe we shouldn't get everybody's Twitter feed in. . . . All we're doing right now is, "Let's just collect more and more and more data and something good will just happen." And that's I think that's kind of wishful thinking.

Law enforcement's adoption of data and analytic tools without a specific technical purpose also surfaced during my time at surveillance industry conferences. When I first observed software representatives interact with potential law enforcement customers, I assumed law enforcement would tell software representatives their needs and ask how the products could help them achieve their operational goals. However, the inverse pattern was more frequently the case: software representatives demonstrated the use of their platform in a non-law enforcement—usually military—context, and then asked local law enforcement whether they would be interested in a similar application in their local context. In other words, instead of filling analytic gaps or technical voids identified by law enforcement, software representatives helped create new kinds of institutional demand.

DISCUSSION: BIG DATA AS SOCIAL

This article draws on unique data to offer an on-the-ground account of big data surveillance. Providing a case study of the Los Angeles Police Department (LAPD), it offers insight into the reasons why the use of big data analytics spread throughout the organization, including factors particular to the LAPD such as consent decree mandates, but also broader isomorphic shifts (DiMaggio and Powell 1983) toward use of predictive analytics across organizational fields. In analyzing how the LAPD uses big data in their surveillance activities, I argue it is both *continuous* and *transformative*: the adoption of advanced analytics facilitates amplifications of existing surveillance practices, but also fundamentally changes daily operations. I describe five key shifts in practice associated with adoption of big data analytics, each of which falls on different points on the continuum between law enforcement and intelligence activities. Whereas the person-based point system and place-based predictive algorithms are largely quantified recapitulations of "traditional" (Marx 2016) surveillance, the inter-institutional integration of data and proliferation of dragnet surveillance practices—including the use of data on individuals with no direct police contact and data gathered from institutions typically not associated with crime control—represent fundamental transformations in the very nature of surveillance.

Big data and associated new technological tools permit unprecedentedly broad and deep surveillance. By broad, I mean surveillance capable of passively tracking a large number of people. Information that would previously have been unknown to law enforcement because it was too labor intensive to retrieve is more readily available, and individuals previously unknown to law enforcement are now part of the corpus through dragnet surveillance and data collection by non–criminal justice organizations. By deep, I mean able to track one individual more intensively over time, including across different institutional settings. The intended and unintended social consequences of new surveillance practices have implications for social inequality, law, and future research on big data surveillance in other fields.

Implications for Social Inequality

The role of the criminal justice system in the reproduction of inequality has received considerable attention in the literature (for a review, see Laub 2014). However, the impact of the use of big data surveillance on inequality remains an open empirical question. The use of new surveillant technologies could either reduce or reinforce existing inequalities. By contributing new insights into how big data plays out on the ground in policing, this research helps adjudicate between the two possibilities.

On the one hand, big data analytics may be a means by which to ameliorate persistent inequalities in policing. Data can be marshaled to replace unparticularized suspicion of racial minorities and human exaggeration of patterns with less biased predictions of risk. Social psychological research demonstrates that humans are “cognitive misers” (Fiske and Taylor 1991) who rely on shortcuts—such as the conflation of blackness and criminality (Quillian and Pager 2001)—to understand the world. Because stereotypes have the most cognitive utility in the face of incomplete information, if big data can be utilized to provide more complete information, it may lead officers to rely less on stereotypes about race and class. In that sense, the use of big data may serve to reduce hyper-surveillance of minority neighborhoods and the consequent erosion of community trust (Sampson and Bartusch 1998). Big data may also be used to “police the police.” Digital trails are susceptible to oversight. Therefore, aggregating data on police practices may shed light on systematic patterns and institutional practices previously dismissed as individual-level bias, ultimately providing an opportunity to increase transparency and accountability. However, transparency and accountability do not flow automatically from big data policing. Data-based surveillance is less visible than traditional street policing methods (Joh 2016) and is embedded in power structures. The outcomes of struggles between law enforcement, civilians, and information technology companies—who increasingly own the storage platforms and proprietary algorithms used in data analysis—will play a role in determining whether big data policing will ameliorate or exacerbate inequalities.

On the other hand, this research highlights how data-driven surveillance practices may be implicated in the reproduction of inequality in at least three ways: by deepening the surveillance of individuals already under suspicion; widening the criminal justice dragnet unequally; and leading people to avoid “surveilling” institutions that are fundamental to social integration. First, mathematized police practices serve to place individuals already

under suspicion under new and deeper forms of surveillance, *while appearing to be objective*, or, in the words of one captain, "just math." Despite the stated intent of the point system to avoid legally contestable bias in police practices, it hides both intentional and unintentional bias in policing and creates a self-perpetuating cycle: if individuals have a high point value, they are under heightened surveillance and therefore have a greater likelihood of being stopped, further increasing their point value. Such practices hinder the ability of individuals already in the criminal justice system from being further drawn into the surveillance net, while obscuring the role of enforcement in shaping risk scores. Moreover, individuals living in low-income, minority areas have a higher probability of their "risk" being quantified than those in more advantaged neighborhoods where the police are not conducting point-driven surveillance. Importantly, this quantified modality of social control has consequences that reach beyond individuals with high point values. Field interview cards record information not only about the individual in question, but also information on people the individual is with. The exponential capture of personal data beyond the primary individuals involved in the police encounter is a strategic means of channeling more individuals into the system, thus facilitating future tracking.

Whereas the point system is consequential for racial and class inequality, if not implemented effectively,[12] place-based algorithms may exacerbate neighborhood inequalities. Historical crime data are incomplete; estimates of unreported crime range from less than 17 percent to over 68 percent, depending on the offense (Langton et al. 2012). Moreover, crime data are not missing at random. Therefore, there is systematic bias in the training data: crimes that take place in public places are more visible to police and therefore more likely to be recorded (Duster 1997); individuals and groups who do not trust the police are less likely to report crimes (Sampson and Bartusch 1998); and police focus their attention and resources on black communities at a disproportionately high rate relative to drug use and crime rates (Beckett et al. 2005). These social dynamics inform the historical crime data that are fed into the predictive policing algorithm. However, once they are inputted as data, the predictions appear impartial; human judgment is hidden in the black box (Pasquale 2014) under a patina of objectivity.

Unchecked predictions may lead to an algorithmic form of confirmation bias, and subsequently, a misallocation of resources. They may justify the over-policing of minority communities and potentially take away resources from individuals and areas invisible to data collection sensors or subject to systematic underreporting. Put differently, the mechanisms for inclusion in criminal justice databases determine the surveillance patterns themselves. Predictive models are performative, creating a feedback loop in which they not only predict events such as crime or police contact, but also contribute to their future occurrence.[13]

Second, new digitized surveillance practices broaden the scope of people law enforcement can track. This can be understood as a new form of "net widening" (Cohen 1985), effectively widening the criminal justice dragnet, and doing so unequally. Consider ALPRs,

12. Place-based algorithms are most effective (and least biased) when predicting crimes with high reporting rates, such as motor vehicle theft.
13. For related work on performativity in a different field—finance—see MacKenzie, Muniesa, and Siu (2007).

one of the primary means of tracking people without police contact. Even though ALPRs are dragnet surveillance tools that collect information on everyone, rather than merely those under suspicion, the likelihood of being inputted into the system is not randomly distributed. Crime and enforcement patterns lead to unequal data capture across individuals, groups, and the city. ALPRs are deployed based on department crime statistics (i.e., to higher crime areas), raising similar questions to those posed earlier about unequal enforcement and reporting practices along lines of race, class, and neighborhood. In that sense, ALPR datasets are investigatory tools for law enforcement, but they are disproportionately "populated by the movements of particular groups" (Renan 2016:1059). Similarly, the ability to build out secondary surveillance networks in Palantir has implications for inequality, as minority individuals and individuals in poor neighborhoods have a higher probability of being in the primary (and thus secondary) surveillance net than do people in neighborhoods where the police are not conducting point-driven or other data-intensive forms of policing.

How are unequal mechanisms for inclusion in the surveillance net consequential for social inequality? Recall the operative theory from a detective that if people are not doing anything wrong, the police should not be looking them up many times over the course of their lives. However, queries are not raw data (Gitelman 2013); rather, they are, in part, a product of enforcement practices. Empirical research consistently demonstrates that stop-and-query patterns are unequally distributed by race, class, and neighborhood (Epp, Maynard-Moody, and Haider-Markel 2014). Quantified practices may thus serve to exacerbate inequalities in stop patterns, create arrest statistics needed to justify stereotypes, and ultimately lead to self-fulfilling statistical prophecies (Merton 1948). Moreover, as police contact is the entry point into the criminal justice system, the digital feedback loops associated with predictive policing may ultimately justify the growth and perpetuation of the carceral state.

One might argue that if you have nothing to hide, being included in police databases is nothing to fear. However, once individuals are in the primary or secondary surveillance net, they can become intelligence targets and linked to future data points. By virtue of being in the system, individuals are more likely—correctly or incorrectly—to be identified as suspicious. Consider how the quantification of previous stops in the point system serves as justification for future stops, or how the detective suggested a man was suspicious because his name had been queried multiple times. Using a series of data points to reconstruct an individual's intentions and behaviors, whether incriminating or exculpatory, rests on the assumption of an infallible state and of actors who run searches without error or prejudice. Much like in DNA databases (Duster 2005; Hindmarsh and Prainsack 2010; Lynch et al. 2008), in order to be a hit, one has to be in the database in the first place. Unequal rates of database inclusion can have real consequences—African Americans are seven times more likely than whites to be wrongly convicted of murder (Gross, Possley, and Stephens 2017). Therefore, analyzing the feeder mechanisms by which individuals are channeled into criminal justice databases helps us better understand how inequalities produced by differential surveillance may be magnified as individuals are processed through the criminal justice system.

Third, integrating external, non-police data into the law enforcement corpus has unanticipated consequences. Although integrated systems create new opportunities for service delivery, they also make surveillance possible across formerly discrete institutional boundaries. By using other institutions' data, criminal justice surveillance practices may have a chilling effect, deterring people from using such institutions and thereby subverting their original mandates. For example, individuals wary of criminal justice surveillance may avoid interacting with important institutions where they would leave a digital trace. Previous research demonstrates that individuals involved in the criminal justice system (i.e., who have been stopped by police, arrested, convicted, or incarcerated) engage in "system avoidance," systematically avoiding surveilling institutions such as medical, financial, educational, and labor market institutions that keep formal records (i.e., put them "in the system") (Brayne 2014). Given that involvement with the criminal justice system is highly stratified, the negative consequences of system avoidance—for future health outcomes, financial self-sufficiency, acquisition of human capital, and upward economic mobility—will be similarly disproportionately distributed, thus exacerbating any preexisting inequalities for an expanding group of already disadvantaged individuals.

This research builds on work on labeling theory, extending the relationship between the stigma of criminal justice contact and inequality into the digital age (Becker 1963; Brayne 2014; Goffman 2014; Goffman 1963; Kohler-Hausmann 2013; Lyon 2006; Pager 2007; Rios 2011; Stuart 2016; Wakefield and Wildeman 2013; Western and Pettit 2005). The integration of records may effectively extend the mark of a criminal record (Pager 2007)—or merely the mark of criminal justice *contact*—into other institutions. This creep of data across institutional contexts can lead to "cascading disadvantages" (Pasquale 2014:218; see also Gandy 2009). As individuals leave more digital traces, a "new economy of moral judgement" (Fourcade and Healy 2017:24) becomes possible. Building on Weber's concept of class situation, Fourcade and Healy (2013) argue that institutions now use actuarial techniques to track, sort, and categorize individuals into "classification situations" with different rewards and punishments attached. These classification situations differentially shape life chances (see also Bowker and Star 2000). For example, classifying individuals as low or high risk for crime, terrorist activity, loan default, or medical conditions structures not only if and how they will be surveilled, but also their life chances more generally. This research begins to account for how the marking process may be changing in the age of digitized policing, and how the big data environment creates potentially farther-reaching digitized collateral consequences of involvement in the criminal justice system.

In summary, the burden of new surveillance practices is not borne equally, nor is the error they produce (Guzik 2009). That said, this research does not necessarily suggest the police intentionally use big data maliciously. Rather, as Barocas and Selbst (2016) argue, discrimination may be, at least in part, an artifact of the data collection and analysis process itself. Algorithmic decision procedures can "reproduce existing patterns of discrimination, inherit the prejudice of prior decision makers, or simply reflect the widespread biases that persist in society" (Barocas and Selbst 2016:674). Understanding each step of data collection and analysis is crucial for understanding how data systems—despite being thought of as objective, quantified, and unbiased—may inherit the bias of their creators and users. As an institution historically implicated in the reproduction of inequality,

understanding the intended and unintended consequences of machine-learned decisions and new surveillant technologies in the criminal justice system is of paramount importance.

Implications for Law

Technological tools for surveillance are far outpacing legal and regulatory responses to the new surveillant landscape. Therefore, the findings from this project have important implications for law. First, current privacy laws—such as the Privacy Act of 1974—are anachronistic because they largely concern controls at the point of data collection. With the increased capacity to store vast amounts of data for significant periods of time, privacy laws now must also account for function creep, protecting individuals from potential future secondary uses of their data. Relatedly, law enforcement routinely purchases privately collected data, blurring the lines between public and private and highlighting the importance of revisiting third-party doctrine in the digital age.[14]

Second, use of big data for predictive analytics challenges the traditional paradigm of Fourth Amendment law, which is *transactional*: it focuses on one-off interactions between law enforcement and a suspect. However, police surveillance is increasingly *programmatic*: it is ongoing, cumulative, and sometimes suspicionless (Renan 2016). In terms of ALPRs, for example, "what begins as more generalized collection can morph into something quite different when the government runs individuated searches in its datasets" (Renan 2016:1053). Therefore, it is an open question whether cumulative surveillance should require different legal frameworks, such as administrative law, "from those that govern each isolated step" (Renan 2016:1058), namely criminal procedure.

Third, when big data—such as predictive policing forecasts—are combined with small data—such as traditional individualized suspicion based on particularized facts about a suspect—it effectively makes it easier for law enforcement to meet the reasonable suspicion standard in practice. In the words of one captain, "Some officer somewhere if this [predictive policing] gets big enough is going to say, 'okay, everybody in the box is open season,' you know? And that's not the case." Therefore, legal scholars such as Ferguson (2015: 336) suggest the courts should "require a higher level of detail and correlation using the insights and capabilities of big data."

Fourth, dragnet surveillance tools such as ALPRs represent a proliferation of pre-warrant surveillance and make everyday mass surveillance possible at an unprecedented scale. Pre-crime data can be mined for links once criminal suspicion comes into play. Once in a database, a suspect can repeatedly be surveilled; law enforcement can retroactively search ALPR data and identify individuals, vehicles, times, and places, rather than starting to gather information on them only once they come under suspicion. The retroactive nature of policing in an era of dragnet data collection means information is routinely accumulated and files are lying in wait. In that sense, *individuals lead incriminating lives*—daily activities, now codified as data, can be marshaled as evidence ex post facto. The proliferation of

14. According to United States v. Miller (1939) and Smith v. Maryland (1979), the third-party doctrine maintains that "when an individual voluntarily shares information with third parties, like telephone companies, banks, or even other individuals, the government can acquire that information from the third-party absent a warrant" (Executive Office of the President 2014).

pre-warrant surveillance tools also creates new opportunities for parallel construction, the process of building a separate evidentiary base for a criminal investigation to conceal how the investigation began, if it involved warrantless surveillance or other inadmissible evidence.

Finally, previous practical constraints, which placed natural limits on the scope of surveillance, are less relevant in light of new dragnet tools. One new analytic technique or surveillant technology on its own may not be consequential, but the combined power of using, for example, the person-based point system in conjunction with ALPR data in conjunction with network diagrams in Palantir grants authorities a level of insight into an individual's life that historically would have constituted a Fourth Amendment search and thus required a warrant. However, because no one of those surveillance practices falls outside the parameters of the law in isolation, neither does their combination.[15] In that sense, intelligence is essentially pre-warrant surveillance. Detectives and prosecutors rarely find a "smoking gun," a member of Palantir's legal counsel explained, but they can now build up a sequence of events that they were previously unable to. By "integrating data into a single ontology," he continued, users can draw connections between actors and depict a coherent scheme. Hunches that would be insufficient grounds for obtaining a warrant can be retroactively backed up using existing data, and queries can be justified in hindsight after data confirm officer suspicions. Instead of needing to justify to a judge why they require a warrant, law enforcement can first take advantage of the surveillance opportunities new technologies provide.

Implications for Research in Other Fields

Big data is being utilized for surveillance practices in a wide range of institutional domains beyond policing, including but not limited to health, finance, credit, marketing, insurance, education, immigration, defense, and activism. Although this study focused on law enforcement, big data surveillance is not something over which the LAPD has exclusive domain. Rather, this case reflects broader institutional shifts toward the use of emergent technologies and advanced analytics. Thinking beyond policing, future research may consider how big data surveillance practices identified in this study may operate in similar or different ways across fields.

Drawing from work in surveillance studies helps us systematically compare use of the newest modality of surveillance—big data—across domains. Informed by Lyon's (2003) theory of surveillance as social sorting and Marx's (2016) discussion of surveillance means, goals, and data attributes, I offer three concrete questions for future research that could help us better understand the use of big data for surveillance across institutional domains: Why was big data surveillance adopted (goals)? How is big data surveillance conducted (means)? What interventions are made based on big data surveillance, and to what consequence (ends)? Table 1 summarizes these questions.

First, why was big data surveillance adopted? What institutional goals was it intended to achieve? Lyon (2003:1) argues that the organizational imperative for surveillance is

[15]See Justice Sotomayor's concurring opinion in United States v. Jones (2012) and Joh (2016).

one of social sorting: “Surveillance today sorts people into categories, assigning worth or risk, in ways that have real effects on their life-chances . . . it is a vital means of sorting populations for discriminatory treatment” (see also Ericson and Haggerty 1997; Rule 2007). He identifies different categories of surveillance, each of which have different mandates and classificatory goals. Actors in the criminal justice system, for example, engage in “categorical suspicion” (Lyon 2003), collecting information to classify individuals according to risk and to identify threats to law and order. The purpose of surveillance in other institutions, however, may not be categorical suspicion but rather “categorical seduction” —classifying customers for targeted marketing, financial services, or credit (Lyon 2003; see also Gandy [1993] on the “panoptic sort”)—or “categorical care”—surveillance in health and welfare organizations aimed at improving services through better coordination of personal data (Ball and Webster 2007). Analyzing changes and continuities in the structure of relationships between agents of surveillance and those who are surveilled (Marx 2016) may help us more fully understand the social process of big data surveillance and its consequences for social stratification.

Whereas surveillance goals have not fundamentally changed much over the past century, surveillance means have transformed considerably. By providing a detailed analysis of how the means of surveillance have changed in the age of big data, this study may inform future research on big data surveillance in other fields. Are the five shifts in practice identified in this study—discretionary to quantified risk assessments, explanatory to predictive analytics, query-based to alert-based systems, moderate to low inclusion thresholds, and disparate to integrated databases—occurring in other institutional contexts? For example, to what extent are data collected beyond their proximate institutional environments being used in healthcare or finance?

The goals and means of big data surveillance inform the final question posed for future research: what are the ends of big data surveillance? What do institutional actors *do* based on insights gleaned from big data surveillance, and with what consequence? Surveillance involves extracting information from different flows (Deleuze and Guattari 1987; Haggerty and Ericson 2000; Marx 2016). Distinct information flows are then reassembled into a “data double”—a digital approximation of individuals based on the electronic traces they leave (Poster 1990:97)—which is used to decide on differential treatment. Digital scores and ranks can be understood as a form of capital (Fourcade and Healy 2017), used to determine who the police stop, who credit bureaus determine as credit-worthy, and who public assistance agencies deem eligible for benefits. Simply put, one’s surveillance profile structures the types of communications, opportunities, constraints, and care one receives. Big data surveillance could therefore have stratifying effects if individuals in positions of structural disadvantage are more likely to be subject to harmful forms of surveillance, and those in positions of structural advantage are more likely to be targeted by advantageous surveillance and classification schemes (Fourcade and Healy 2017). Therefore, categorical suspicion, seduction, and care may have very different implications for social inequality, depending on what institutional actors *do* based on the intelligence acquired through big data surveillance.

This article demonstrates that in the digital age, individuals leave data traces hundreds of times throughout the day, each of which contributes to the corpus of big data that a growing number of institutions use for decision-making. Institutional actors making decisions based on big data may assume that data doubles are more accurate, or unbiased, representations of a person's profile than are those gleaned from "small" data, such as personal observations. However, this perspective obscures the social side of big data surveillance. Systematic bias —whether intentional or unintentional—exists in training data used for machine learning algorithms, and it may be an artifact of human discretion or the data mining process itself. Moreover, the implications of false positives and false negatives associated with big data surveillance vary widely across domains. The stakes for being wrongly arrested for a crime you did not commit are very different from receiving a movie recommendation not to your taste. Furthermore, categories of surveillance are not mutually exclusive in the age of big data, as information can be shared across previously separate institutional boundaries. For example, electronic medical records were originally created to improve prescription drug and care coordination, but they are increasingly used to police the illicit use and sale of prescription drugs. The places where categories of surveillance intersect, and therefore have ambiguous implications for inequality, may be particularly fruitful sites for future sociological inquiry.

Finally, future research may examine the political economy underpinning the procurement of analytic software that organizations use for big data surveillance. Examining the genealogies of surveillance technologies, for example, reveals that many of the resources for developing big data analytics come from federal funds. In the law enforcement context, those grants quickly become subsumed into police organizations' operating budgets. Therefore, departments have an incentive to continue using big data—or appear to be using it—even if it is not an effective means of solving the organization's first-order problems, such as reducing crime.

Understanding the implications of big data surveillance is more complex than simply knowing who is surveilled more or less. Instead, we need to understand who is surveilled by whom, in what way, and for what purpose. How surveillance structures life chances may differ according to the goals, means, and ends involved. Although surveillance is a generalizable organizational imperative, big data is changing the means of surveillance. Accordingly, this article helps us better understand *how* big data surveillance is conducted, and calls for systematic research on the relationship between the goals, means, and ends of big data surveillance across institutional domains.

CONCLUSIONS

Through a case study of the Los Angeles Police Department, this article analyzed the role of big data in surveillance practices. By socially situating big data, I examined why it was adopted, how it is used, and what the implications of its use are. Focusing on the interplay between surveillance practices, law, and technology offers new insights into social control and inequality. I argued that big data participates in and reflects existing social structures. Far from eliminating human discretion and bias, big data represents a new form of capital that is both a social product and a social resource. What data law enforcement collects,

their methods for analyzing and interpreting it, and the way it informs their practice are all part of a fundamentally social process. Characterizing predictive models as "just math," and fetishizing computation as an objective process, obscures the social side of algorithmic decision-making. Individuals' interpretation of data occurs in preexisting institutional, legal, and social settings, and it is through that interpretive process that power dynamics come into play.

Use of big data has the potential to ameliorate discriminatory practices, but these findings suggest implementation is of paramount importance. As organizational theory and literature from science and technology studies suggests, when new technology is overlaid onto an old organizational structure, long-standing problems shape themselves to the contours of the new technology, and new unintended consequences are generated. The process of transforming individual actions into "objective" data raises fundamentally sociological questions that this research only begins to address. In many ways, it transposes classic concerns from the sociology of quantification about simplification, decontextualization, and the privileging of measurable complex social phenomena onto the big data landscape.

Surveillance is always ambiguous; it is implicated in both social inclusion and exclusion, and it creates both opportunities and constraints. The way in which surveillance helps achieve organizational goals and structure life chances may differ according to the individuals and institutions involved. Examining the means of big data surveillance across institutional domains is an open and timely line of inquiry, because once a new technology is disseminated in an institutional setting, it is difficult to scale back.

Acknowledgments

I wish to thank Devah Pager for her invaluable support and guidance. I am also grateful for the many helpful comments I received from Paul DiMaggio, Janet Vertesi, Kim Lane Scheppele, Maria Abascal, David Pedulla, Becky Pettit, the Social Media Collective at Microsoft Research, and the *ASR* editors and anonymous reviewers. Finally, I wish to thank anonymous individuals within the Los Angeles Police Department for making this research possible.

Funding

This research was funded by the Horowitz Foundation for Social Policy. Additional support was provided by grant, 5 R24 HD042849, awarded to the Population Research Center at The University of Texas at Austin by the Eunice Kennedy Shriver National Institute of Child Health and Human Development.

Biography

Sarah Brayne is an Assistant Professor of Sociology and Faculty Research Associate in the Population Research Center at the University of Texas at Austin. Using qualitative and quantitative methods, her research examines the use of "big data" within the criminal justice system as well as the consequences of surveillance for law and social inequality.

References

Angwin Julia. 2014. Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance. New York: Times Books.

Ball Kirstie, and Webster Frank, eds. 2007. The Intensification of Surveillance: Crime, Terrorism & Warfare in the Information Age. London, UK: Pluto Press.

Barley Stephen R. 1986. “Technology as an Occasion for Structuring: Evidence from Observations of CT Scanners and the Social Order of Radiology Departments.” Administrative Science Quarterly 31(1):78–108. [PubMed: 10281188]

Barley Stephen R. 1996. “Technicians in the Workplace: Ethnographic Evidence for Bringing Work into Organization Studies.” Administrative Science Quarterly 41(3):404–441.

Barocas Solon, and Selbst Andrew D.. 2016. “Big Data’s Disparate Impact.” California Law Review 104:671–732.

Becker Howard S. 1963. Outsiders: Studies in the Sociology of Deviance. New York: Free Press.

Beckett Katherine, Nyrop Kris, Pfingst Lori, and Bowen Melissa. 2005. “Drug Use, Drug Possession Arrests, and the Question of Race: Lessons from Seattle.” Social Problems (52)3:419–41.

Bittner Egon. 1967. “The Police on Skid-Row: A Study of Peace Keeping.” American Sociological Review 32(5):699–715.

Bonczar Thomas P., and Herberman Erinn J.. 2014. “Probation and Parole in the United States, 2013.” Washington, DC: Bureau of Justice Statistics.

Bowker Geoffrey C., and Star Susan Leigh. 2000. Sorting Things Out: Classification and Its Consequences. Cambridge, MA: MIT Press.

boyd danah, and Crawford Kate. 2012. “Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon.” Information, Communication & Society 15(5):662–79.

Braga Anthony A., and Weisburd David L.. 2010. Policing Problem Places: Crime Hot Spots and Effective Prevention. New York: Oxford University Press.

Braverman Harry. 1974. Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century. New York: Monthly Review Press.

Brayne Sarah. 2014. “Surveillance and System Avoidance: Criminal Justice Contact and Institutional Attachment.” American Sociological Review 79(3):367–91.

Brown v. Plata. 2011. 563 U.S. 493.

Browne Simone. 2015. Dark Matters: On the Surveillance of Blackness. Durham, NC: Duke University Press.

Carson E Ann. 2015. “Prisoners in 2014.” Washington, DC: Bureau of Justice Statistics.

Christin Angèle. 2016. “From Daguerreotypes to Algorithms: Machines, Expertise, and Three Forms of Objectivity.” ACM Computers & Society 46(1):27–32.

Cohen Stanley. 1985. Visions of Social Control: Crime, Punishment and Classification. Malden, MA: Polity Press.

Deleuze Gilles, and Guattari Felix. 1987. A Thousand Plateaus: Capitalism and Schizophrenia. Minneapolis: The University of Minnesota Press.

DiMaggio Paul J., and Powell Walter W.. 1983. “The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields.” American Sociological Review 48(2):147–60.

Duster Troy. 1997. “Pattern, Purpose and Race in the Drug War.” Pp. 206–287 in Crack in America: Demon Drugs and Social Justice, edited by Reinarman C and Levine HG. Berkeley: University of California Press.

Duster Troy. 2005. “Race and Reification in Science.” Science 307:1050–51. [PubMed: 15718453]

Epp Charles R., Maynard-Moody Steven, and Haider-Markel Donald P.. 2014. Pulled Over: How Police Stops Define Race and Citizenship. Chicago: University of Chicago Press.

Ericson Richard V., and Haggerty Kevin D.. 1997. Policing the Risk Society. Toronto: University of Toronto Press.

Ericson Richard V., and Haggerty Kevin D.. 2006. The New Politics of Surveillance and Visibility. Toronto: University of Toronto Press.

Espeland Wendy N., and Vannebo Berit I.. 2007. “Accountability, Quantification and Law.” Annual Review of Law and Society 3:21–43.

Executive Office of the President. 2014. “Big Data: Seizing Opportunities, Preserving Values.” Washington, DC: The White House.

Feeley Malcolm M., and Simon Jonathan. 1992. “The New Penology: Notes on the Emerging Strategy of Corrections and Its Implications.” Criminology 30(4):449–74.

Ferguson Andrew G. 2015. “Big Data and Predictive Reasonable Suspicion.” University of Pennsylvania Law Review 63(2):327–410.

Fiske John. 1998. “Surveilling the City: Whiteness, the Black Man and Democratic Totalitarianism.” Theory, Culture and Society 15(2):67–88.

Fiske Susan T., and Taylor Shelley E.. 1991. Social Cognition, 2nd ed. New York: McGraw-Hill.

Foucault Michel. 1977. Discipline and Punish: The Birth of the Prison. New York: Random House.

Fourcade Marion, and Healy Kieran. 2013. “Classification Situations: Life-Chances in the Neoliberal Era.” Accounting, Organizations and Society 38:559–72.

Fourcade Marion, and Healy Kieran. 2017. “Seeing like a Market.” Socioeconomic Review 15(1):9–29.

Gandy Oscar H. 1993. The Panoptic Sort: A Political Economy of Personal Information. Boulder, CO: Westview Press.

Gandy Oscar H. 2002. “Data Mining and Surveillance in the Post-9.11 Environment.” Presentation to the Political Economy Section, International Association for Media and Communication Research.

Gandy Oscar H. 2009. Coming to Terms with Chance: Engaging Rational Discrimination and Cumulative Disadvantage. Farnham, UK: Ashgate Publishing.

Garland David. 2001. The Culture of Control: Crime and Social Order in Contemporary Society. New York: Oxford University Press.

Giddens Anthony. 1990. The Consequences of Modernity. Stanford, CA: Stanford University Press.

Gilliom John. 2001. Overseers of the Poor: Surveillance, Resistance, and the Limits of Privacy. Chicago: University of Chicago Press.

Gitelman Lisa, ed. 2013. Raw Data Is an Oxymoron. Cambridge, MA: MIT Press.

Goffman Alice. 2014. On the Run: Fugitive Life in an American City. Chicago: University of Chicago Press.

Goffman Erving. 1963. Stigma: Notes on the Management of Spoiled Identity. Englewood Cliffs, NJ: Prentice-Hall.

Gross Samuel R., Possley Maurice, and Stephens Kalara. 2017. “Race and Wrongful Convictions in the United States.” National Registry of Exonerations, Newkirk Center for Science and Society, University of California-Irvine.

Gustafson Kaaryn S. 2011. Cheating Welfare: Public Assistance and the Criminalization of Poverty. New York: NYU Press.

Guzik Keith. 2009. “Discrimination by Design: Predictive Data Mining as Security Practice in the United States’ ‘War on Terrorism.’” Surveillance and Society 7(1):1–17.

Hacking Ian. 1990. The Taming of Chance. Cambridge, UK: Cambridge University Press.

Haggerty Kevin D., and Ericson Richard V.. 2000. “The Surveillant Assemblage.” British Journal of Sociology 51(4):605–622. [PubMed: 11140886]

Harcourt Bernard E. 2006. Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age. Chicago: University of Chicago Press.

Hindmarsh Richard, and Prainsack Barbara, eds. 2010. Genetic Suspects: Global Governance of Forensic DNA Profiling and Databasing. Cambridge, UK: Cambridge University Press.

Innes Martin. 2001. “Control Creep.” Sociological Research Online 6:1–10.

Joh Elizabeth E. 2016. “The New Surveillance Discretion: Automated Suspicion, Big Data, and Policing.” Harvard Law and Policy Review 10(1):15–42.

Kitchin Rob. 2014. The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences. London, UK: SAGE.

Kling Rob. 1991. “Computerization and Social Transformations.” Science, Technology & Human Values 16(3):342–67.

Kohler-Hausmann Issa. 2013. “Misdemeanor Justice: Control without Conviction.” American Journal of Sociology 119(2):351–93.

Laney Doug. 2001. “3D Data Management: Controlling Data Volume, Velocity, and Variety.” Stamford, CT: META Group.

Langton Lynn, Berzofsky Marcus, Krebs Christopher, and Smiley-McDonald Hope. 2012. “Victimizations Not Reported to the Police, 2006–2010.” Washington, DC: Bureau of Justice Statistics.

Laub John H. 2014. “Understanding Inequality and the Justice System Response: Charting a New Way Forward.” New York: William T. Grant Foundation.

Lazer David, and Radford Jason. 2017. “Data ex Machina: Introduction to Big Data.” Annual Review of Sociology 43:19–39.

Los Angeles Police Department. 2017. “Sworn Personnel by Rank, Gender, and Ethnicity (SPRGE) Report” (http://www.lapdonline.org/sworn_and_civilian_report).

Lynch Michael, Cole Simon A., McNally Ruth, and Jordan Kathleen. 2008. Truth Machine: The Contentious History of DNA Fingerprinting. Chicago: University of Chicago Press.

Lyon David. 1994. The Electronic Eye: The Rise of Surveillance Society. Minneapolis: University of Minnesota Press.

Lyon David, ed. 2003. Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination. New York: Routledge.

Lyon David, ed. 2006. Theorizing Surveillance: The Panopticon and Beyond. New York: Polity.

Lyon David. 2015. Surveillance after Snowden. New York: Polity.

MacKenzie Donald, Muniesa Fabian, and Siu Lucia, eds. 2007. Do Economists Make Markets? On the Performativity of Economics. Princeton, NJ: Princeton University Press.

Manning Peter K. 2011. The Technology of Policing: Crime Mapping, Information Technology, and the Rationality of Crime Control. New York: NYU Press.

Manning Peter K., and Van Maanen John, eds. 1978. Policing: A View from the Street. New York: Random House.

Marx Gary T. 1974. “Thoughts on a Neglected Category of Social Movement Participant: The Agent Provocateur and the Informant.” American Journal of Sociology 80(2):402–442.

Marx Gary T. 1988. Undercover: Police Surveillance in America. Berkeley: University of California Press.

Marx Gary T. 1998. “Ethics for the New Surveillance.” The Information Society 14(3):171–85.

Marx Gary T. 2002. “What’s New About the ‘New Surveillance’? Classifying for Change and Continuity.” Surveillance and Society 1(1):9–29.

Marx Gary T. 2016. Windows into the Soul: Surveillance and Society in an Age of High Technology. Chicago: University of Chicago Press.

Mayer-Schönberger Viktor, and Cukier Kenneth. 2013. Big Data: A Revolution That Will Transform How We Live, Work, and Think. New York: Houghton Mifflin Harcourt.

Merton Robert K. 1948. “The Self-Fulfilling Prophecy.” The Antioch Review 8(2):193–210.

Meyer John W., and Rowan Brian. 1977. “Institutionalized Organizations: Formal Structure as Myth and Ceremony.” American Journal of Sociology 83(2):340–63.

Mohler George O., Short Martin B., Malinowski Sean, Johnson Mark, Tita George E., Bertozzi Andrea L., and Brantingham P. Jeffrey. 2015. “Randomized Controlled Field Trials of Predictive Policing.” Journal of the American Statistical Association 110:1399–1411.

Monahan Torin, and Palmer Neal A.. 2009. “The Emerging Politics of DHS Fusion Centers.” Security Dialogue 40(6):617–36.

Moskos Peter. 2008. Cop in the Hood: My Year Policing Baltimore’s Eastern District. Princeton, NJ: Princeton University Press.

Oxford American Dictionary of Current English. 1999. Oxford: Oxford University Press.

Pager Devah. 2007. Marked: Race, Crime, and Finding Work in an Era of Mass Incarceration. Chicago: University of Chicago Press.

Papachristos Andrew V., Hureau David M., and Braga Anthony A.. 2013. “The Corner and the Crew: The Influence of Geography and Social Networks on Gang Violence.” American Sociological Review 78(3):417–47.

Pasquale Frank. 2014. The Black Box Society: The Secret Algorithms That Control Money and Information. Cambridge, MA: Harvard University Press.

Perry Walter L., Brian McInnis Carter C. Price, Smith Susan C., and Hollywood John S.. 2013. "Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations." RAND Safety and Justice Program. Santa Monica, CA. Retrieved July 6, 2017 (http://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf).

Porter Theodore M. 1995. Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton, NJ: Princeton University Press.

Poster Mark. 1990. The Mode of Information. Chicago: University of Chicago Press.

Quillian Lincoln, and Pager Devah. 2001. "Black Neighbors, Higher Crime? The Role of Racial Stereotypes in Evaluations of Neighborhood Crime." American Journal of Sociology 107(3):717–67.

Ratcliffe Jerry. 2008. Intelligence-Led Policing. Cullompton, UK: Willan Publishing.

Reiss Albert J. 1971. The Police and the Public. New Haven, CT: Yale University Press.

Renan Daphna. 2016. "The Fourth Amendment as Administrative Governance." Stanford Law Review 68(5):1039–1129.

Rios Victor M. 2011. Punished: Policing the Lives of Black and Latino Boys. New York: NYU Press.

Roush Craig R. 2012. "Quis Custodiet Ipsos Custodes? Limits on Widespread Surveillance and Intelligence Gathering by Local Law Enforcement after 9/11." Marquette Law Review 96(1):315–76.

Rule James B. 1974. Private Lives and Public Surveillance: Social Control in the Computer Age. New York: Schocken Books.

Rule James B. 2007. Privacy in Peril: How We Are Sacrificing a Fundamental Right in Exchange for Security and Convenience. New York: Oxford University Press.

Sampson Robert J., and Bartusch Dawn J.. 1998. "Legal Cynicism and (Subcultural?) Tolerance of Deviance: The Neighborhood Context of Racial Differences." Law and Society Review 32:777–804.

Scott Richard W. 1987. Organizations: Rational, Natural, and Open Systems. Englewood Cliffs, NJ: Prentice-Hall.

Scott Richard W. 2004. "Reflections on a Half-Century of Organizational Sociology." Annual Review of Sociology 30:1–21.

Sherman Lawrence W. 2013. "Targeting, Testing and Tracking Police Services: The Rise of Evidence-Based Policing, 1975–2025." Crime and Justice 42(1):377–451.

Sherman Lawrence W., Gartin Patrick R., and Buerger Michael E.. 1989. "Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place." Criminology 27(1):27–55.

Skogan Wesley G. 2006. Police and Community in Chicago: A Tale of Three Cities. New York: Oxford University Press.

Smith v. Maryland. 1979. 442 U.S. 735.

Soss Joe, Fording Richard C., and Schram Sanford F.. 2011. Disciplining the Poor: Neoliberal Paternalism and the Persistent Power of Race. Chicago: University of Chicago Press.

Stuart Forrest. 2016. Down, Out & Under Arrest: Policing and Everyday Life in Skid Row. Chicago: University of Chicago Press.

Terry v. Ohio. 1968. 392 U.S. 1.

Tracy Paul E., and Morgan Vincent. 2000. "Big Brother and His Science Kit: DNA Databases for 21st Century Crime Control." Journal of Criminal Law and Criminology 90(2):635–90.

Travis Jeremy, Western Bruce, and Redburn Steve, eds. 2014. Growth of Incarceration in the United States: Exploring Causes and Consequences. Washington, DC: National Academy of Science.

Uchida Craig D., and Swatt Marc L.. 2015. "Operation LASER and the Effectiveness of Hotspot Patrol: A Panel Analysis." Police Quarterly 16(3):287–304.

United States v. Jones. 2012. 132 S. Ct. 945, 565 U.S.

United States v. Miller. 1939. 307 U.S. 174.

U.S. Department of Justice. 2001 [2015]. “L.A. Consent Decree.” Washington, DC: U.S. Department of Justice.

Wakefield Sara, and Uggen Christopher. 2010. “Incarceration and Stratification.” Annual Review of Sociology 36:387–406.

Wakefield Sara, and Wildeman Christopher. 2013. Children of the Prison Boom: Mass Incarceration and the Future of American Inequality. New York: Oxford University Press.

Waxman Matthew C. 2009. “Police and National Security: American Local Law Enforcement and Counter-Terrorism after 9/11.” Journal of National Security Law and Policy 3:377–407.

Weber Max. 1978. Economy and Society: An Outline of Interpretive Sociology. Berkeley: University of California Press.

Weisburd David, Mastrofski Stephen D., Ann Marie McNally Rosann Greenspan, and Willis James J.. 2003. “Reforming to Preserve: COMPSTAT and Strategic Problem-Solving in American Policing.” Criminology and Public Policy 2:421–56.

Western Bruce. 2006. Punishment and Inequality in America. New York: Russell Sage Foundation.

Western Bruce, and Pettit Becky. 2005. “Black-White Wage Inequality, Employment Rates, and Incarceration.” American Journal of Sociology 111(2):553–78.

White House Police Data Initiative. 2015. “Launching the Police Data Initiative” (https://obamawhitehouse.archives.gov/blog/2015/05/18/launching-police-data-initiative).

Whren v. United States. 1996. 517 U.S. 806.

Willis James J., Mastrofski Stephen D., and Weisburd David. 2007. “Making Sense of COMPSTAT: A Theory-Based Analysis of Organizational Change in Three Police Departments.” Law and Society Review 41(1):147–88.

Wilson James Q. 1968. Varieties of Police Behavior: The Management of Law and Order in Eight Communities. Cambridge, MA: Harvard University Press.

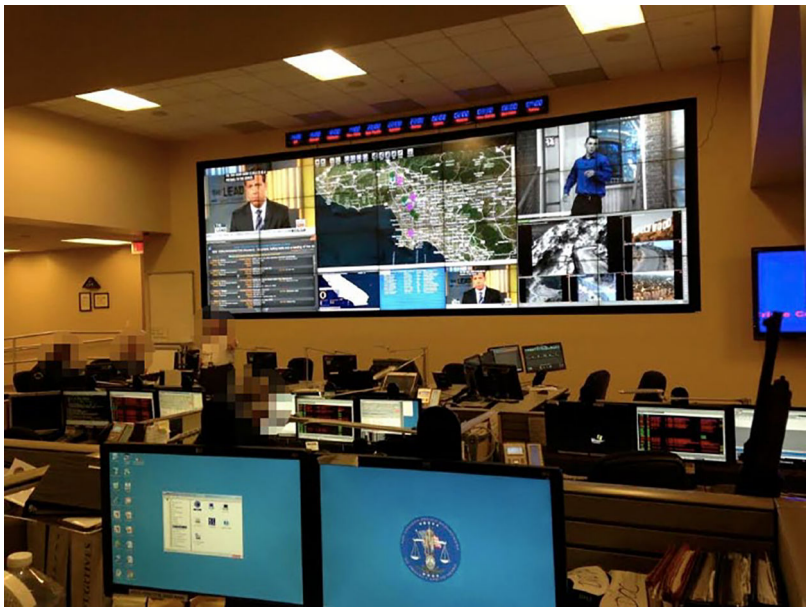

Figure 1.
Situation Room at the Real-Time Crime Analysis Center (RACR)
Source: Author's photo.

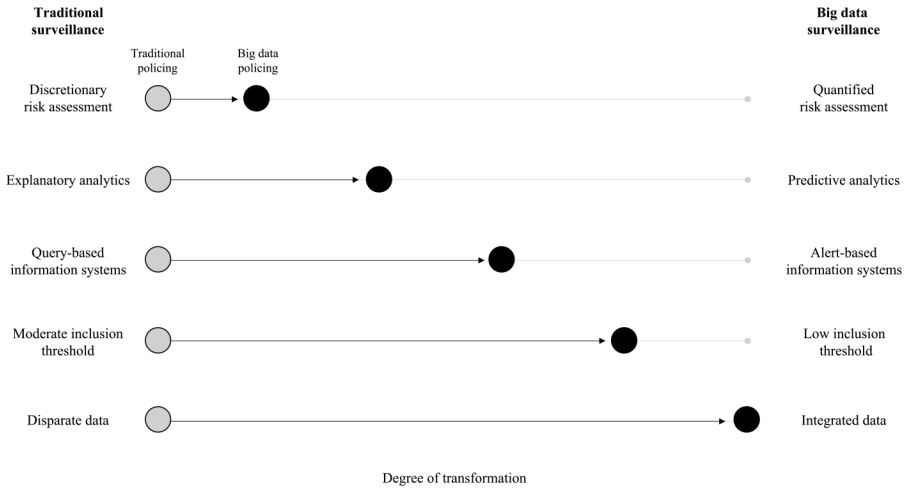

Figure 2.
Migration of Traditional Police Practices toward Big Data Surveillance

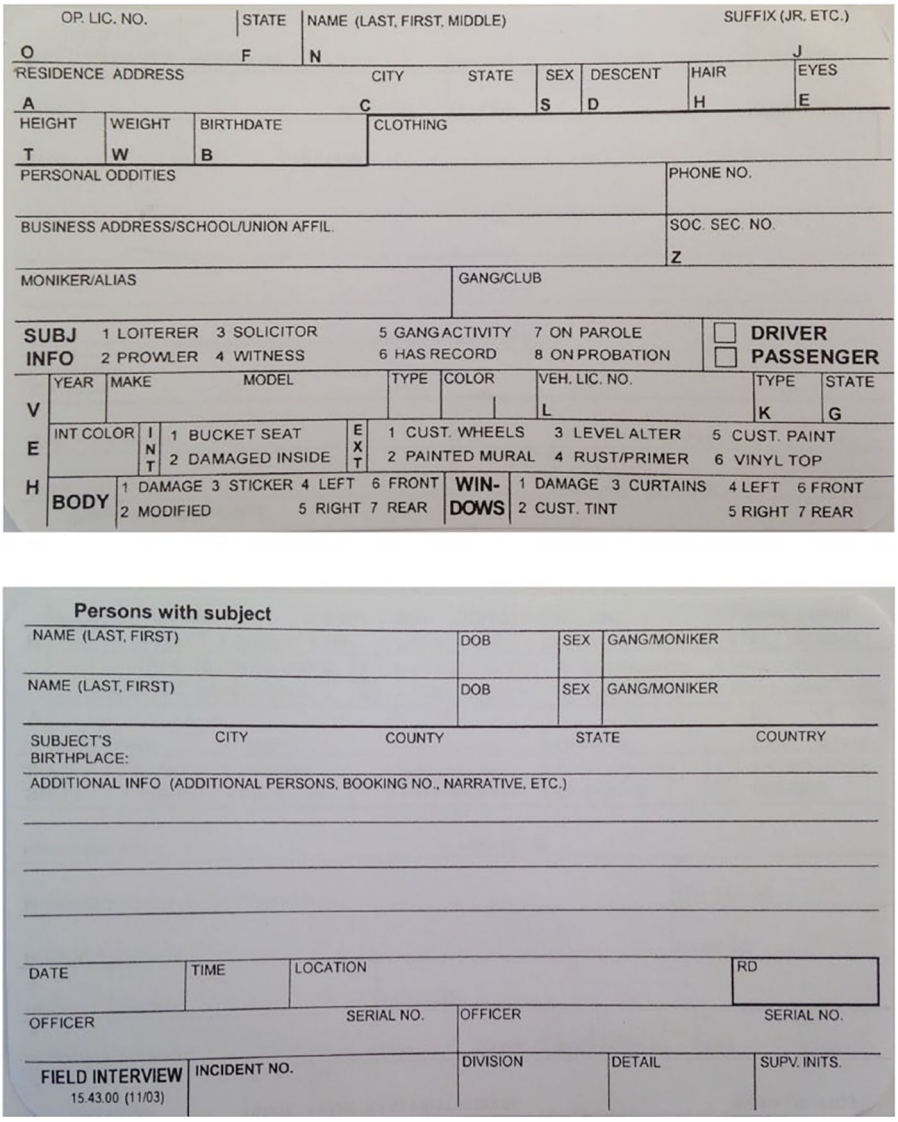

Figure 3.
Field Interview (FI) Cards
Source: LAPD.

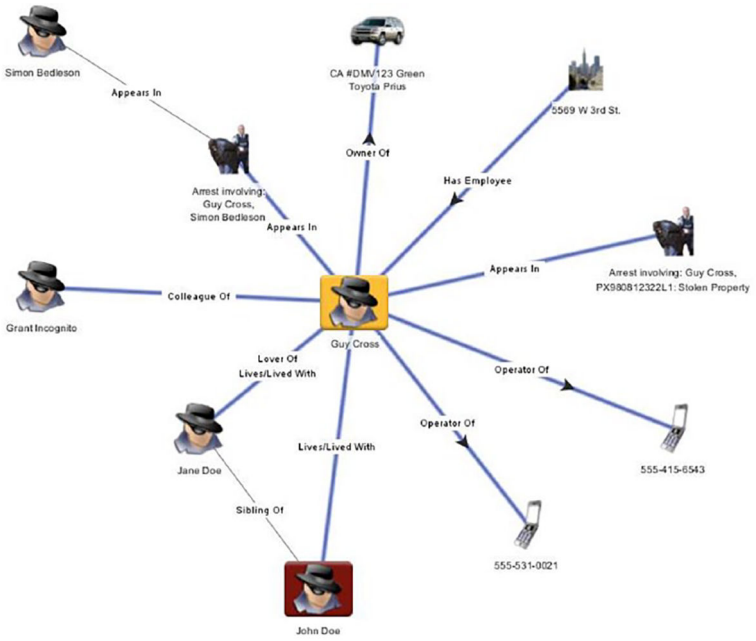

Figure 4.
Network in Palantir
Source: Palantir Technologies.

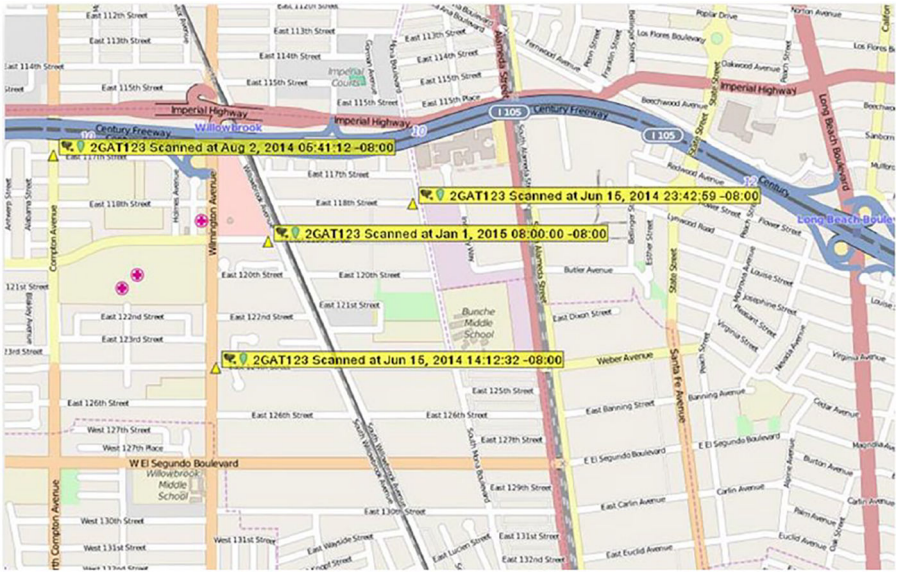

Figure 5.
Plotted ALPR Readings
Source: Palantir Technologies.

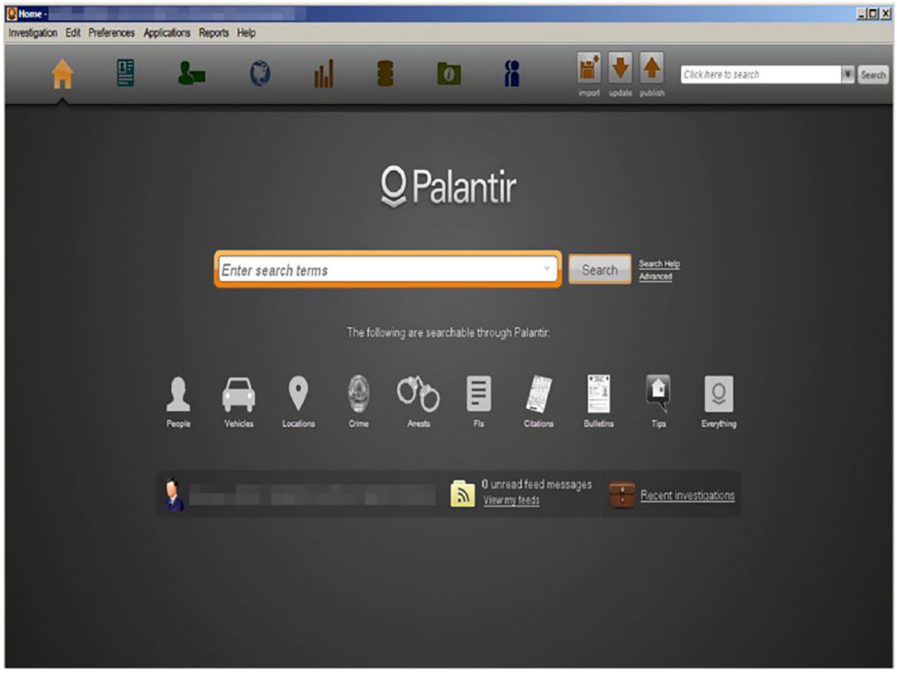

Figure 6.
Palantir Homepage
Source: LAPD.

Table 1.

Framework for Analyzing Big Data Surveillance across Institutional Contexts

	Goals		Means	Ends	
Types of Surveillance	**Institutional Field**	**Relationship between Individual and Institution**	**Shifts in Surveillance Practices Associated with Big Data**	**Institutional Interventions**	**Consequences for Inequality**
Categorical Suspicion	Criminal justice, intelligence	Classifying individuals according to risk; potential as criminals/ terrorists	1) Discretionary to quantified risk assessment 2) Explanatory to predictive analytics	Marking, apprehension, social control	Stigma, spillover into other institutions
Categorical Seduction	Finance, marketing, credit	Classifying individuals according to their value to companies; potential as customers	3) Query-based to alert-based systems 4)Moderate to low inclusion thresholds 5) Disparate to integrated data	Different products, perks, access to credit, opportunities, constraints	Upward or downward economic mobility; reproducing current patterns
Categorical Care	Medical care, public assistance	Classifying individuals according to their need; potential as clients		Personalized medicine, welfarist service delivery	May reduce inequality except when intersects with suspicion or seduction

Am Sociol Rev. Author manuscript; available in PMC 2024 February 06.