The Rise of Big Data Policing

Surveillance, Race, and the Future of Law Enforcement

Andrew Guthrie Ferguson

NEW YORK UNIVERSITY PRESS
New York

Introduction

Big Data Policing

A towering wall of computer screens blinks alive with crisis. A digital map of Los Angeles alerts to 911 calls. Television screens track breaking news stories. Surveillance cameras monitor the streets. Rows of networked computers link analysts and police officers to a wealth of law enforcement intelligence. Real-time crime data comes in. Real-time police deployments go out. This high-tech command center in downtown Los Angeles forecasts the future of policing in America.[1]

Welcome to the Los Angeles Police Department's Real-Time Analysis Critical Response (RACR) Division. The RACR Division, in partnership with Palantir—a private technology company that began developing social network software to track terrorists—has jumped head first into the big data age of policing.[2]

Just as in the hunt for international terror networks, Palantir's software system integrates, analyzes, and shares otherwise-hidden clues from a multitude of ordinary law enforcement data sources. A detective investigating a robbery suspect types a first name and a physical description into the computer—two fragmented clues that would have remained paper scraps of unusable data in an earlier era.[3] The database searches for possible suspects. Age, description, address, tattoos, gang affiliations, vehicle ownership instantly pop up in sortable fields. By matching known attributes, the computer narrows the search to a few choices. A photograph of a possible suspect's car is identified from an automated license-plate reader scouring the city for data about the location of every vehicle. Detectives follow up with a witness to identify the car used in the robbery. A match leads to an arrest and a closed case.[4]

A 911 call. A possible gang fight in progress. RACR Command directs patrol units to the scene all the while monitoring their real-time progress. Data about the fight is pushed to officers on their mobile phones.[5]

Alerts about past shootings and gang tensions warn officers of unseen dangers.[6] Neighborhood histories get mapped for insight. Officers scroll through photographs to visualize the physical geography before they arrive. All of the data is instantaneously sent to officers, allowing them to see the risks before they need to act.[7]

Roll call. Monday morning. Patrol officers receive digital maps of today's "crime forecast."[8] Small red boxes signify areas of predicted crime. These boxes represent algorithmic forecasts of heightened criminal activity: years of accumulated crime data crunched by powerful computers to target precise city blocks. Informed by the data, "predictive policing" patrols will give additional attention to these "hot" areas during the shift.[9] Every day, police wait in the predicted locations looking for the forecast crime. The theory: put police in the box at the right time and stop a crime. The goal: to deter the criminal actors from victimizing that location.

Soon, real-time facial-recognition software will link existing video surveillance cameras and massive biometric databases to automatically identify people with open warrants.[10] Soon, social media feeds will alert police to imminent violence from rival gangs.[11] Soon, data-matching technologies will find suspicious activity from billions of otherwise-anonymous consumer transactions and personal communications.[12] By digitizing faces, communications, and patterns, police will instantly and accurately be able to investigate billions of all-too-human clues.

This is the future. This is the present. This is the beginning of big data policing.[13]

Big data technologies and predictive analytics will revolutionize policing.[14] Predictive policing, intelligence-driven prosecution, "heat lists" of targets, social media scraping, data mining, and a data-driven surveillance state provide the first clues to how the future of law enforcement will evolve.

At the center of policing's future is data: crime data, personal data, gang data, associational data, locational data, environmental data, and a growing web of sensor and surveillance sources. This big data arises from the expanded ability to collect, store, sort, and analyze digital clues about crime.[15] Crime statistics are mined for patterns, and victims of violence are mapped in social networks. While video cameras watch our movements, private consumer data brokers map our interests and sell that information to law enforcement.[16] Phone numbers, emails, and finances can all be studied for suspicious links. Government agencies collect health, educational, and criminal records.[17] Detectives monitor public Facebook, YouTube, and Twitter feeds.[18] Aggregating data centers sort and study the accumulated information in local and federally funded fusion centers.[19] This is the big data world of law enforcement—still largely in its infancy but offering vastly more incriminating bits of data to use and study.

Behind the data is technology: algorithms, network analysis, data mining, machine learning, and a host of computer technologies being refined and improved every day. Police can identify the street corner most likely to see the next car theft[20] or the people most likely to be shot.[21] Prosecutors can target the crime networks most likely to destabilize communities,[22] while analysts can link suspicious behaviors for further investigation.[23] The decisional work of identifying criminal actors, networks, and patterns now starts with powerful computers crunching large data sets almost instantaneously. Math provides the muscle to prevent and prosecute crime.

Underneath the data and technology are people—individuals living their lives. Some of these people engage in crime, some not. Some live in poverty, some not. But all now find themselves encircled by big data's reach. The math behind big data policing targets crime, but in many cities, crime suppression targets communities of color. Data-driven policing means aggressive police presence, surveillance, and perceived harassment in those communities. Each data point translates to real human experience, and many times those experiences remain fraught with all-too-human bias, fear, distrust, and racial tension. For those communities, especially poor communities of color, these data-collection efforts cast a dark shadow on the future.

This book shines light on the "black data" arising from big data policing:[24] "black" as in opaque, because the data exists largely hidden within complex algorithms; "black" as in racially coded, because the data directly impacts communities of color; "black" as in the next new thing, given legitimacy and prominence due to the perception that data-driven anything is cool, techno-friendly, and futuristic; and, finally, "black" as distorting, creating legal shadows and constitutional gaps where the law used to see clearly. Black data matters because it has real-world impacts.

Black data marks human “threats” with permanent digital suspicion and targets poor communities of color. Black data leads to aggressive use of police force, including deadly force, and new forms of invasive surveillance. Big data policing, and these new forms of surveillance and social control, must confront this black data problem.

This book examines how big data policing impacts the “who,” “where,” “when,” and “how” of policing. New technologies threaten to impact all aspects of policing, and studying the resulting distortions provides a framework to evaluate all future surveillance technologies. A race is on to transform policing. New developments in consumer data collection have merged with law enforcement’s desire to embrace “smart policing” principles in an effort to increase efficiency amid decreasing budgets. Data-driven technology offers a double win—do more with less resources, and do so in a seemingly objective and neutral manner.

This book arises out of the intersection of two cultural shifts in policing. First, predictive analytics, social network theory, and data-mining technology have all developed to a point of sophistication such that big data policing is no longer a futuristic idea. Although police have long collected information about suspects, now this data can be stored in usable and sharable databases, allowing for greater surveillance potential. Whereas in an earlier era a police officer might see a suspicious man on the street and have no context about his past or future danger, soon digitized facial-recognition technologies will identify him, crime data will detail his criminal history, algorithms will rate his risk level, and a host of citywide surveillance images will provide context in the form of video surveillance for his actions over the past few hours. Big data will illuminate the darkness of suspicion. But it also will expand the lens of who can be watched.

The second cultural shift in policing involves the need to respond to outrage arising from police killings of unarmed African Americans in Ferguson, Missouri; Staten Island, New York; Baltimore, Maryland; Cleveland, Ohio; Charleston, South Carolina; Baton Rouge, Louisiana; Falcon Heights, Minnesota; and other cities. This sustained national protest against police—and the birth of the Movement for Black Lives—brought to the surface decades of frustration about racially discriminatory law enforcement practices.[25] Cities exploded in rage over unaccountable police actions. In response, data-driven policing began to be sold as one answer to racially discriminatory policing, offering a seemingly race-neutral, "objective" justification for police targeting of poor communities.[26] Despite the charge that police data remains tainted by systemic bias,[27] police administrators can justify continued aggressive police practices using data-driven metrics. Predictive policing systems offer a way seemingly to turn the page on past abuses, while still legitimizing existing practices.

For that reason, my aim in this book is to look at the dangers of black data arising at this moment in history. Only by understanding why the current big data policing systems were created and how traditional policing practices fit within those systems can society evaluate the promise of this new approach to data-driven law enforcement. Black data must be illuminated to see how it might be abused. The promise of "smarter" law enforcement is unquestionably real, but so is the fear of totalizing surveillance. Growing "law and order" rhetoric can lead to surveillance overreach. Police administrators, advocates, communities, and governments must confront those concerns before—not after—the technology's implementation. And society must confront those challenges informed by an understanding of how race has fractured and delegitimized the criminal justice system for many citizens. Black data, of course, is not just about African Americans, although the history of racially discriminatory policing runs deep in certain communities. But black data exposes how all marginalized communities face a growing threat from big data policing systems. People of color, immigrants, religious minorities, the poor, protesters, government critics, and many others who encounter aggressive police surveillance are at increased risk. But so is everyone, because every one of us produces a detailed data trail that exposes personal details. This data—suctioned up, sold, and surveilled—can be wrong. The algorithmic correlations can be wrong. And if police act on that inaccurate data, lives and liberty can be lost.

Big data is not all dystopian. The insights of big data policing need not be limited to targeting criminal activity. The power of predictive analytics can also be used to identify police misconduct or identify the underlying social and economic needs that lead to crime. In an era of heighted concern with police accountability, new surveillance technologies offer new avenues to watch, monitor, and even predict police misconduct. Systems of "blue data" can be created to help "police the police."

Similarly, big data technologies can be redirected to identify and target social, economic, or environmental risk factors. This is the promise of "bright data," in which the surveillance architecture developed to police criminal risk can be redirected to address environmental risks and social needs. After all, just because big data policing identifies the risk, this does not mean that law enforcement must provide the remedy.

The big data policing revolution has arrived. The singular insight of this innovation is that data-driven predictive technologies can identify and forecast risk for the future. Risk identification is also the goal of this book—to forecast the potential problems of big data policing as it reshapes law enforcement. Long-standing tensions surrounding race, secrecy, privacy, power, and freedom are given new life in digital form with the advent of big data analytics. New technologies will open up new opportunities for investigation and surveillance. The technological environment is rich with possibility but also danger. This book seeks to initiate a conversation on the growth of these innovations, with the hope that by exposing and explaining the distorting effects of data-driven policing, society can plan for its big data future.

1

Big Data's Watchful Eye

The Rise of Data Surveillance

The world is full of obvious things which nobody by any chance ever observes.
—Sherlock Holmes[1]

Data Trails

You are being watched. Surveilled. Tracked. Targeted. Every search on the internet recorded.[2] Every purchase at the store documented.[3] Every place you travel mapped.[4] They know how fast you drive, your preferred cereal, your dress size. They know your financial situation, all of your past jobs, your credit limit.[5] They know your health concerns, reading preferences, and political voting patterns. They also know your secrets. They have been watching for years.[6] In truth, you live in a surveillance state. The watchers know you because of the data you leave behind.

But it is not just you. These watchers also know about your family, friends, neighbors, colleagues, clubs, and associates. They see the circles you contact, the friends you ignore, and the political issues you embrace. They see you as part of a group, but they also see all the other parts of the group.[7] Links expand outward, so that all of your contacts can be visualized as a web of interrelated, interconnected groups.

Welcome to the world of big data, where one's data trail reveals the mosaic of lived experience and has become the currency of a new economy.[8] "They" are companies, companies that enable a digital world by offering convenience, information, and services all in return for one thing: data. Your personal data and interests—all of those points of commercial interaction, consumer choice, "likes," links, and loves—have been vacuumed up, processed, and sold to others wanting to get to know you. Currently, this widespread surveillance remains in the hands of for-profit companies, for the purpose of offering consumers convenience and choice. But law enforcement is interested too.[9] And most of this information is a subpoena (or warrant) away from being part of a criminal case. The investigative lure of big data technologies is just too powerful to ignore.

What Is Big Data?

To understand the potential of big data policing, the scope of big data must be explored. So what is big data? In general, "big data" is a shorthand term for the collection and analysis of large data sets with the goal to reveal hidden patterns or insights.[10] A report from the Executive Office of the President summarized: "There are many definitions of 'big data' which may differ depending on whether you are a computer scientist, a financial analyst, or an entrepreneur pitching an idea to a venture capitalist. Most definitions reflect the growing technological ability to capture, aggregate, and process an ever-greater volume, velocity, and variety of data."[11] In simple terms, large collections of data can be sorted by powerful computers to visualize unexpected connections or correlations.[12] Machine-learning tools and predictive analytics allow educated guesses about what the correlations mean.[13]

A simple example of how big data works can be seen at Amazon.com. Beneath each item for sale is a recommendation section that displays information about what "customers who bought this item also bought" and items that are "frequently bought together." Amazon generates these suggestions from the purchasing patterns of its 300 million customers who bought related items. Correlating the historical data of billions of transactions leads to an insight into which goods customers usually purchase together. Amazon, of course, also knows everything you have ever bought from the company. But Amazon can sort the purchasing data of any particular product to show the consumer patterns of all past customers. Amazon can use that large data set to predict what items you might actually want in the future.[14] After all, if you bought a coffee maker today, you may need coffee tomorrow.

A more unusual example involves the correlation between Pop-Tarts and hurricanes. Walmart—a company that collects more than two and half petabytes of data every hour from customers[15] (equivalent to 50 million four-drawer filing cabinets filled with text)—discovered that just before a hurricane, people buy an unusual amount of Strawberry Pop-Tarts.[16] Why? No one really knows. Perhaps the reason for the uptick is because Pop-Tarts are nonperishable comfort food, and sometimes sugary comfort is just what you need after a big storm. Or perhaps not. Big data demonstrates the correlation, not the cause. It offers insight without explanation—a reality that is both useful and unsettling.

Obviously, big companies like Amazon and Walmart collect personal data, but what is the extent of big data collection across our daily lives? More than can be comprehended. As Julia Angwin termed it, "We are living in a Dragnet Nation—a world of indiscriminate tracking where institutions are stockpiling data about individuals at an unprecedented pace."[17] The World Privacy Forum—a watchdog group on personal privacy—estimates that there are 4,000 different databases collecting information on us.[18] Every time we interact with computers, sensors, smartphones, credit cards, electronics, and much, much more, we leave a digital trail that is revealing of ourselves and valuable to others.[19] These are the breadcrumbs of the big data maze. Follow them and they lead right back to you.

Where Does Big Data Come From?

Big data comes from you. You provide the building blocks of big data's power in small digital bits.

Think of the normal patterns of your life. You probably live in a house or an apartment. Even if you do not live in a wired "smart home" that comes equipped with a "smart fridge" to order milk when you run out, or a Nest "smart thermostat" to turn down the heat when you leave, your home does reveal basic data about your lifestyle.[20] You have an address. The address reveals general information about your income (as implied by the cost of the home) and your family size (number of bedrooms). Your zip code provides clues about demographics, wealth, and political sentiment.

You probably get mail at that address. First to note, the United States Postal System runs the Mail Isolation Control and Tracking program, which photographs the exterior of every single piece of mail processed in the United States.[21] So data about your address is tracked along with the 150 billion letters mailed each year.[22] But more obviously, your mail also reveals things about you. Magazine subscriptions reveal your political and cultural interests, and catalogues reveal your hobbies and shopping preferences. Mail reveals your friends and associates, just as packages reveal your styles, interests, and lifestyle choices. Even junk mail says something about what marketers think you want.

You likely also use the internet. Some of those packages came from online shopping. Those online retail companies track your purchases and even those things you looked at but did not purchase.[23] Inferences from those purchases are also valuable. If you bought infant diapers for the first time, you might also need age-appropriate children's toys for the next holiday season (and for the next 18 years). If you bought a "how to quit smoking book," you might not be the best candidate for a new cigar magazine. But you don't even have to shop to give up your data. Google records every internet search, really every click of the mouse.[24] That means every health query, travel question, childrearing tip, news article, and entertainment site. Google and other search engines provide little windows into your thinking (if not your soul). Your internet protocol (IP) provides your exact location,[25] and while your IP addresses might change as you switch from your home computer to your iPhone to your work computer, Hotmail knows where you are at all times. Amazon knows the page you stopped reading on your Kindle ebook reader.[26] Your cable provider (which may also be your cellphone and wireless provider) knows what TV shows you watch late at night. Netflix and other streaming entertainment services rely on personalized predictive formulas based on past viewing data.

Social media expands the web of data from yourself to your friends and associates.[27] On Facebook, you literally display your "likes" of certain things. Professional sites like LinkedIn add more information about what you do, who you know, and what accolades you have received. Personal and social updates broadcast life changes, and charity or community service activities get promoted. Photos provide data about where you have been and who you were with. Geotagging of information from those photos and other services reveal the time, location, and date of the picture.[28] Facial recognition links people together, so that your photos (and thus your identity) can be tracked over different social media platforms. And sometimes you might simply tell people on Twitter what you are doing or upload photos of your dinner entrée on Instagram or Snapchat.

You might leave your home in a car—a car registered to your address with a name, gender, birthdate, and identification number. The car can be tracked through a city via surveillance cameras, electronic toll collectors, or automated license-plate scanners.[29] Your type of car (hybrid or Hummer) might reveal a lifestyle preference or environmental worldview. The car itself might have Global Positioning System (GPS) tracking through something like a GM OnStar program to allow for instant help in an accident or emergency.[30] But that helpful service requires constant locational tracking. Or maybe you have an insurance provider that monitors real-time driving data of your habits in return for lower car-insurance rates.[31] You drive carefully, you save money.

But, no matter, if you possess a smartphone with locational services turned on, the speed, location, and direction of your car is being monitored in real time.[32] Your iPhone knows a wealth of locational information about where you go, which health clinic you stopped at, and the Alcoholics Anonymous meeting you just attended. Locational data from Google Maps tracks your church attendance, political protests, and friends. Other mobile apps leech data to companies in return for targeted advertisements or travel tips.[33] Games, services, geotracking ads, emergency calls—all depend on location. Everything that little pocket computer does can be tracked and recorded in granular detail. That means that every YouTube video, every photograph, and every check of the weather is collected, to reveal the things you do on a daily basis, as well as where you were when you did them.[34]

Maybe you took that car to work. Your employment history has been harvested by credit agencies.[35] Your job, finances, professional history, and even your education are recorded.[36] Maybe you went shopping. That customer-loyalty card offering in-store discounts also tracks each purchase you make.[37] Stores know not only everything you have purchased going back years but also your physical location when you made the purchase. Maybe you went to the bank. All of your financial information, account balances, late fees, investments, credit history—all are recorded.[38] Your credit card statement is a little reminder of everything you did and where you did it for the past month. Maybe you took the car to have fun. The Google search of local restaurant reviews followed by a map search of a particular restaurant and an Open Table reservation provide a pretty good prediction of your Saturday-night plans.[39]

If you add in "smart devices" connected through the Internet of Things (Fitbits, smart bandages, smart cups) or sensors built into our transportation infrastructure, clothing, and bodies, you have a very revealing web of data about our activities.[40] Researchers predict that there will be over 50 billion smart things connected among the "Internet of Everything" by 2020.[41] These "smart devices" are scarily aware of you. If your television responds to your voice or your electronic personal assistant answers your questions, it means these smart devices are always listening and always on.

Finally, public records filled with census data, property records, licenses, Department of Motor Vehicle information, bankruptcies, criminal convictions, and civil judgments can be purchased by companies seeking to understand us.[42] This official, bureaucratic record of life, linked as it is to governmental data systems, has become the foundation for many credit histories and personalized data dossiers on individuals.[43]

This is how big data becomes big. This is why big data can be such a threat to privacy, associational freedom, and autonomy. Your self-surveillance provides the currency for commercial profit but also the building blocks for an intrusive police state. Every digital clue—with the appropriate legal process—can be demanded by police and prosecutors. Whereas in an earlier era, only your family might know what you did, what you ate, how you dressed, or what you thought about, now the digital clues of life online can be collected, reassembled, and mapped to mirror this same knowledge. In fact, your digital clues may reveal secrets you have kept hidden from your spouse, family, or best friends.

Who Owns the Data?

Private data brokers collect, buy, and sell personal data to companies interested in selling products, determining financial credit risk, or conducting employment background investigations.[44] Data brokers sell your data to others—including law enforcement—for investigative purposes.[45]

Data brokers challenge conventional assumptions about individual privacy. Aggregated private transactions are repurposed and repackaged into a composite targeted profile of you as a consumer.[46] The United States Senate Commerce Committee detailed how big data companies like Acxiom claim to have information on over 700 million consumers worldwide with over 3,000 data segments for nearly every U.S. consumer.[47] Another company, Datalogix, claims to have data on almost every U.S. household.[48] Much of this information is demographic, such as name, address, telephone number, email, gender, age, marital status, children, educational level, and political affiliation. Some of the information is available through consumer transactions, detailing where one bought something, and some of the information focuses on health problems and medical data. The Senate report detailed how "one company collects data on whether consumers suffer from particular ailments, including Attention Deficit Hyperactivity Disorder, anxiety, depression, diabetes, high blood pressure, insomnia, and osteoporosis, among others; another keeps data on the weights of individuals in a household."[49] And "an additional company offers for sale lists of consumers under 44 different categories of health conditions, including obesity, Parkinson's disease, Multiple Sclerosis, Alzheimer's disease, and cancer, among others."[50]

The level of detail can be remarkably creepy.[51] Here are two excerpts from the Senate Commerce Committee's report:

> Equifax maintains approximately 75,000 individual data elements for its use in creating marketing products, including information as specific as whether a consumer purchased a particular soft drink or shampoo product in the last six months, uses laxatives or yeast infection products, OB/GYN doctor visits within the last 12 months, miles traveled in the last 4 weeks, and the number of whiskey drinks consumed in the past 30 days.[52]

> Some companies offer "data dictionaries" that include more than one thousand potential data elements, including whether the individual or household is a pet owner, smokes, has a propensity to purchase prescriptions through the mail, donates to charitable causes, is active military or a veteran, holds certain insurance products including burial insurance or juvenile life insurance, enjoys reading romance novels, or is a hunter.[53]

The companies know if you have allergies, if you smoke or wear contacts, if your elderly parents live with you, if you speak Spanish, the type of roof on your house, and if you have more than 250 Twitter followers.[54] The creepiness crosses into almost comedic stereotypes as large groups of people become lumped together on the basis of shared demographics or income. Data brokers segment out certain groups. Single men and women over age 66 with "low educational attainment and low net worths" are targeted as "Rural Everlasting."[55] Other singles in the same age group but with more disposable income are known as "Thrifty Elders." Certain low-income minority groups composed of African Americans and Latinos are labeled as "Urban Scramble" or "Mobile Mixers."[56] Private data companies regularly sell and repackage this information about consumer activity to other data brokers, further expanding the webs of shared data.

If you think about what big data companies do in the consumer space, you will see the allure for law enforcement. Data brokers collect personal information to monitor individuals' interests and inclinations. They investigate connections among groups of like-minded people and uncover patterns in the data to reveal hidden insights. This is also what law enforcement investigators do with criminal suspects and gangs. Police monitor, investigate, uncover, and target. Police look for suspicious patterns. Police watch. The tools of big data are the tools of surveillance, and law enforcement relies on surveillance to solve and prevent crime. Unsurprisingly, police have shown great interest in the possibilities of big data policing.

A Permanent Digital Record

The first step in solving any crime is analyzing the clues. Knowing who might be the likely suspect has been part of policing since the mid-1700s, when courts first recorded those who were thought to have been involved in a fraud or felony.[57] Unsurprisingly, as policing developed in sophistication, so did data collection and use. The modern "police blotter" now sits on a cloud server accessible to officers across the jurisdiction or the country.[58]

Federal databases like the National Crime Information Center (NCIC) contain 13 million active records, all searchable by police officers on the street or in their patrol cars. In a routine traffic stop, if a police officer "runs your name" through the system, NCIC will provide personal details about any arrests, warrants, gang affiliations, terrorism ties, supervised release, or fugitive status, as well as information about property including gun ownership, car and boat licenses, and even if you have been the victim of identity theft.[59] This massive database filled with state, local, and federal information is reportedly accessed 12 million times *a day* by authorities.[60] The federal government also maintains watch lists focused on terrorism, including 700,000 names in the Terrorist Screening Database (TSD), a million names in the Terrorist Identities Datamart Environment (TIDE), and 50,000 names on the "No-Fly List."[61]

States also collect and generate data sets to monitor citizens. Eleven states maintain extensive electronic gang databases on suspected gang members.[62] Over 800,000 men and women are listed in federal and state sex-offender registries for convicted sex offenders.[63] Individuals convicted of gun crimes in some states have been required to register.[64] Details about where these offenders live, work, and go to school; what cars they drive; and even their appearance (tattoos, facial hair, scars) are updated constantly in digital archives.[65]

After the terrorist attacks of September 11, 2001, federal and state officials joined forces to establish a national intelligence strategy to improve criminal justice data collection and information sharing.[66] A vast array of law enforcement organizations now share personal data about suspects, crimes, and crime patterns. These organizations include state, local, tribal, and territorial agencies, the Department of Justice (DOJ), the Department of Homeland Security (DHS), the Federal Bureau of Investigation (FBI), the Drug Enforcement Administration (DEA), and the Bureau of Alcohol, Tobacco, and Firearms (ATF). A network of fusion centers seeks to share threat-related information across federal and state lines.[67] Regional Information Sharing Systems (RISS) Centers coordinate incoming data, while Crime Analysis Centers (CACs) analyze collected data. These new data-sharing entities also coordinate with the 17 different agencies that make up the United States intelligence community, including outward, international-facing data-collection agencies like the National Security Agency (NSA) and the Central Intelligence Agency (CIA).

Data projects like the National Data Exchange Program (N-DEx) have been set up "as a giant data warehouse" to pull together otherwise-incompatible police databases.[68] As described in N-DEx's "Privacy Impact Assessment,"

> N-DEx provides a national investigative information sharing system available through a secure Internet site that allows criminal justice agencies to search and analyze data representing the entire criminal justice cycle, including crime incident and investigation records; arrest, booking, and incarceration records; and probation and parole records. As a repository of information from local, state, regional, tribal, and federal criminal justice entities, N-DEx provides these agencies with the capability to make linkages between crime incidents, criminal investigations, and related events to help solve, deter, and prevent crimes. . . . N-DEx contains the personally identifiable information (PII) of suspects, perpetrators, witnesses and victims, and anyone else who may be identified in a law enforcement report concerning a crime incident or criminal investigation.[69]

As of 2014, N-DEx had over 107,000 users and over 170 million searchable records.[70] Start-up companies have been building similar private data-management systems to assist law enforcement in organizing the ever-growing stores of data.

Beyond investigative records, law enforcement now collects biological data. Biometric collection regularly includes DNA, fingerprints, photographs, and iris and retina scans—all secured in searchable databases to investigate crimes.[71] The Combined DNA Index System (CODIS) includes 12 million searchable DNA profiles.[72] The FBI's Next Generation Identification (NGI) system integrates fingerprints, palm prints, facial recognition, and iris scans in one larger searchable database. The FBI has over 23 million searchable photographs and the largest collection of fingerprints in the world.[73] All of this data pushes police investigation into the future, and all of it opens the opportunity for new big data tools to sort, search, and discover otherwise hidden connections between crime and criminals.

Data has also revolutionized how certain police run their day-to-day operations. Many large police departments follow the crime numbers to guide strategy.[74] Some bigger police departments like the New York Police Department (NYPD) have gone so far as to hire a director of analytics to assist in crunching the numbers.[75] Other police departments have partnered with private data-analytics companies or consultants to sort and study the collected data. Professional crime analysts routinely participate in strategy sessions in big police departments.[76] While relying on data differently, most have accepted the underlying principle that the big data technologies created for the private sector can assist police administrators working to improve public safety. In fact, in 2009, Los Angeles Police Department (LAPD) chief Charlie Beck wrote a seminal article, titled "Predictive Policing: What Can We Learn from Wal-Mart and Amazon about Fighting Crime in a Recession?," which explicitly advocated adopting data-driven business principles to improve policing.[77] "Analytics," "risk-based deployment," "prediction," "data mining," and "cost-effectiveness" all emerged as new values and goals for the modern police professional.

Currently, consumer big data technologies and law enforcement data systems operate on separate tracks. What Google knows is not what the FBI knows. The NCIC system is not available to private data brokers. A patchwork of federal privacy laws theoretically restricts the direct governmental collection of personal identifiable information. These statutes include the Privacy Act of 1974,[78] Electronic Communications Privacy Act of 1986 (ECPA),[79] and Stored Communications Act (SCA),[80] Foreign Intelligence Surveillance Act (FISA),[81] E-Government Act of 2002,[82] Financial Privacy Act,[83] Communications Act, Gramm-Leach-Bliley Act,[84] Bank Secrecy Act,[85] Right To Financial Privacy Act,[86] Fair Credit Reporting Act,[87] Health Insurance Portability and Accountability Act of 1996 (HIPAA),[88] Genetic Information Non-discrimination Act (GINA),[89] Children's Online Privacy Protection Act (COPPA),[90] Family Educational Rights and Privacy Act,[91] Telephone Records and Privacy Protection Act of 2006,[92] and Video Protection Privacy Act.[93] In addition to being dated (since some were drafted decades before the big data era), these laws do not prevent law enforcement access. As Erin Murphy has written, "The United States Code currently contains over twenty separate statutes that restrict both the acquisition and release of covered information. . . . Yet across this remarkable diversity, there is one feature that all these statutes share in common: each contains a provision exempting law enforcement from its general terms."[94] Police can obtain certain data with a court order or subpoena.[95] With a valid warrant, police can obtain most anything big data companies have collected for consumer purposes. This patchwork of privacy law also does not stop law enforcement from purchasing the same big data information like any other customer.[96] Just like a private data broker, police can purchase your cellphone and internet information directly from the companies.[97]

A complete big data convergence between private consumer data collection and public law enforcement collection has not yet occurred. But the lines are blurry and growing fainter. Once data has been collected in one place, it becomes harder and harder not to aggregate the information. Private data becomes part of public records, and then public records become the building blocks of private and government databases. Data gets sold and repackaged such that the original collection point becomes obscured.[98] If police want to know about a suspect, and the data has been collected by private third parties, those private companies are hard-pressed to push back and protect the information from lawful government requests. A few powerful technology companies have on occasion rejected government requests for assistance in obtaining customer information or have designed encrypted systems to avoid being in a position to provide information to police investigators.[99] But other companies have been good corporate citizens and provided the information as requested.

Big Data Tools

What big data knows is one thing, but the technology used to manipulate and organize that data is the bigger thing.

The real promise of big data remains with the ability to sort, study, and target within large data sets.[100] Big data becomes intelligible because of algorithms and the large-scale computer-processing power now available. Algorithms are just mathematical processes established to solve a particular task. Using algorithms, pattern-matching tools can flag abnormal financial patterns; social network technologies can link groups via emails, addresses, or any common variable; and predictive analytics can take data-driven insights and forecast future events. Machine-learning algorithms powered by artificial intelligence models can sort vast streams of data in ways unimaginable in earlier eras. Collectively, these math tools allow data analysts to divine insight from an otherwise overwhelming amount of information.[101]

As an example of one such insight, the retail giant Target figured out a way to predict when women are pregnant.[102] By studying women who signed up for an in-store baby registry, Target noticed that these self-identified pregnant women shared a similar, repeating purchasing pattern. Pregnant women would purchase folic acid and vitamin supplements in the first trimester (to improve prenatal health), unscented lotion in the second trimester (due to heightened olfactory sensitivity), and hand sanitizer close to their due dates (to protect the newborn from germs). So now if any woman's purchases follow that pattern (even if she has not signed up for a baby registry), Target flags her as pregnant.[103] The correlation of three unrelated consumer purchases leads to a very personal future prediction.

Big data policing is no different. Law enforcement can identify drug dealers from patterns of supplies (purchasing tiny ziplock bags, rubber bands, digital scales), suspicious transactions (depositing cash, high-end all-cash purchases), and travel patterns (to and from a source city for drugs). The information does not need to be 100% accurate (just as sometimes you receive the wrong catalogue in the mail), but—the theory goes—better information allows police to prioritize and target the higher risks to a community. As Cathy O'Neil wrote in her book *Weapons of Math Destruction*, just as Amazon uses data to identify the "recidivist" shopper, police can use data to predict the future criminal.[104]

Big data tools create the potential for big data policing. The combination of new data sources, better algorithms, expanding systems of shared networks, and the possibility of proactively finding hidden insights and clues about crime has led to a new age of potential surveillance. Instead of consumer surveillance, the goal of big data policing is criminal surveillance.

Chapter 2 looks at why police administrators have been open to big data's embrace. Technology has not been the only force pushing innovation. Budget cuts after a national financial recession forced police to change.[105] In addition, long-standing fissures in police/community relations widened as complaints of racial bias and unconstitutional policing grew louder.[106] Protests challenged continued police violence. Communities demanded change from systemic practices of social control like aggressive stop-and-frisks. Out of this frustration, the seemingly objective metrics of data-driven policing became quite appealing. Turning the page on human bias or racial discrimination became an important spur in the adoption of big data policing. The next chapter explores the lure of these new technologies to solve age-old policing problems.