KNOWLEDGE CREATION, DISSEMINATION, AND PRESERVATION STUDIES

Wood, Stacy. 2018. The Paradox of Police Data. *KULA: knowledge creation, dissemination, and preservation studies* 2(1): 9. DOI: https://doi.org/10.5334/kula.34

RESEARCH ARTICLE

The Paradox of Police Data

Stacy Wood
University of Pittsburgh, US
sewood@pitt.edu

This paper considers the history and politics of 'police data.' Police data, I contend, is a category of endangered data reliant on voluntary and inconsistent reporting by law enforcement agencies; it is also inconsistently described and routinely housed in systems that were not designed with long-term strategies for data preservation, curation or management in mind. Moreover, whereas US law enforcement agencies have, for over a century, produced and published a great deal of data about crime, data about the ways in which police officers spend their time and make decisions about resources—as well as information about patterns of individual officer behavior, use of force, and in-custody deaths—is difficult to find. This presents a paradoxical situation wherein vast stores of extant data are completely inaccessible to the public. This paradoxical state is not new, but the continuation of a long history co-constituted by technologies, epistemologies and context.

Keywords: police; critical data studies; platforms; open data; public data; public records

The Paradox of Police Data

Calls for the free and open exchange of government information are not new, but conversations around open data have recently taken on a new valence.[1] Over the past decade, critics calling for open data initiatives at all levels of government have become emboldened (Arcidiacono and Reale 2017; Milan and van der Velden 2016), seeing such measures as a potential boon to transparency and accountability as well as a mechanism for citizen participation, a reconceptualization of data as a public good, and the generation of secondary economic value through public data reuse. Open data projects such as the nationally focused and temporarily federally supported Police Data Initiative, as well as citizen-generated initiatives such as Campaign Zero, expose the complex and unique case of police data within this context of a shifting culture around public data. They also simultaneously point to a long history of attempts at gathering data about crime, police and the activities of policing (Smith and Austin Jr. 2015; Campaign Zero 2015). In assembling a history of federal police data collection policies and practices, this paper argues that policing itself has been transformed by its own collection and analysis of data, from the initial attempts to document crime at the turn of the century to the permeation of routinized, structured reporting to the contemporary implementation of cloud-based evidence management systems and attempts to harness data for prediction and risk management. Each of these eras and their representative approaches to police data collection are co-constituted by the emergence of new technologies alongside cultural and philosophical orientations toward crime and criminality, communicating not only how information gathering occurs but also what the relationship between information and change, reform or abolition might be.

Daily, massive amounts of data are generated not only through the interactions between police and citizens but also through the interactions between police and their use of technology. The majority of this data remains unavailable to the public and housed and analyzed in proprietary systems, creating an interesting paradox in which data on policing remains a scarce resource even as it proliferates. Even further, the category of police data is misleading as the majority of data produced and published by law enforcement

[1] For more on the history of open data movements, see Jocelyn Cranefield, Oliver Robertson, and Gillian Oliver, 2014, 'Value in the Mash: Exploring the Benefits, Barriers and Enablers of Open Data Apps'; Tim Davies, 2010, 'Open Data, Democracy and Public Sector Reform'; and Jeffrey Alan Johnson, 2014, 'From Open Data to Information Justice.'

agencies is not about police officers, but rather about crime. Data about the ways in which police officers spend their time and make decisions about resources, as well as about patterns of individual officer behavior, use of force, and in-custody deaths, is difficult to find. This paper asserts that so-called police data, even in its abundance, is a category of endangered data reliant on the interest of law enforcement agencies, described to varying degrees and housed in a range of systems rarely tasked with long-term strategies for data preservation, curation or management. Police data is one of many categories of data supplied by the United States government that is essentially preserved through reporting rather than through a sophisticated data preservation strategy, prescribing reuse and interpretation and challenging traditional modes of practice around records management and archival management. This essay provides a brief history of federal police data collection requirements in the United States. An overview of the literature on police data collection offers a framework for understanding the relationship between prevailing attitudes toward crime and emergent technologies deployed by police. In the final section, I outline five potential consequences of the contemporary reliance on proprietary records management systems connected in the cloud.

The Uniform Crime Report and the Origins of 'Police Data' Collection

The Uniform Crime Report (UCR), established in 1927, was the first national crime data system in the United States. It emerged out of a partnership between the International Association for Police Chiefs (IACP) and the Social Science Research Council (SSRC) as a first attempt to gather and compare statistics from individual police agencies across the country. Previous to this effort, attempts to gather data about crime in the United States focused largely on documenting the flows of the incarcerated population for the 1850 United States Census (Rosen 1995). Early data gathering throughout the nineteenth century reflected a reckoning with shifting understandings of and institutionalized responses to crime and criminality. By 1820, in the early days of American penitentiaries, crime was typically understood as a problem with the morality and character of the individual (Barnes 1921). The prevailing models for organizing the lives of inmates within the penitentiary focused on a combination of repentance and punishment, and although the Pennsylvania system that championed wholesale isolation and religious contemplation for inmates eventually faded away in favor of the Auburn system, which privileged regimentation and surveillance, each foregrounded the individual (Meskell 1999). By the end of the nineteenth century, however, crime and criminality were seen as both symptomatic of larger social phenomena and as something hereditary, as characteristic of groups rather than individuals (Jenkins 1984). Organizations like the National Congress of Penitentiary and Reformatory Discipline in Cincinnati, Ohio (formed in 1870) understood crime as a collective moral disease (Wines 1871). Alongside the rise of eugenics and phrenology, rhetoric surrounding crime and criminality was highly racialized and mobilized within penitentiaries through reform organizations that treated the penitentiaries like laboratories (Stoler 1989). As the Progressive Era took hold at the end of the nineteenth century, new theories followed that attributed criminality to outgrowths of rapid urbanization, rapid flows of immigration and the deterioration of traditional support systems and structures. The gathering and management of information became endemic to the management, control and improvement of human behavior and collective well-being (Crunden 1982).

The early incarnation of the UCR organized data collection around seven crimes considered fundamental for comparing crime rates—murder, forcible rape, burglary, aggravated assault, larceny and motor vehicle theft. On the whole, the UCR relied on self-reporting from police departments themselves and remained relatively stagnant. The initial report, released in January of 1930, reported data from four hundred cities throughout the United States. It is worth noting that these numbers have not drastically increased; in the latest preliminary semi-annual Uniform Crime Report published by the FBI, statistics are available for four hundred and seventy-four cities throughout the United States (Federal Bureau of Investigation [FBI] 2017). Also in 1930, Congress passed legislation giving the office of the Attorney General the ability to 'acquire, collect, classify, and preserve identification, criminal identification, crime and other records,' as well as the ability to delegate and secure appointments for those in charge of these data collection initiatives (28 U.S. Code § 534). The Attorney General then designated the FBI as the clearinghouse for the UCR; the FBI took over UCR activities in September 1930 and still administers the report to this day.

The methods used to collect data for the UCR have continued to evolve over the decades. In 1979, an eighth category, arson, was added to the crimes tracked by the UCR through a congressional directive (FBI 2011). Data collection changed even more significantly after the 1985 publication of a new 'Blueprint for the Future of the Uniform Crime Reporting Program' (Poggio et al.). This document analyzed the successes and failures of the previous fifty years and outlined several solutions to the extant problems in the system. Perhaps most significantly, the report suggested scrapping the previous method of collecting data on eight

types of crime and replacing it with a system that divided crime reporting into two *classes*. These classes would later become known as 'Part I Index Crimes' (made up of the original eight serious crimes) and 'Part II Index Crimes' (made up of 21 less commonly reported crimes). Today, law enforcement agencies report the number of known crimes once a month to the UCR.

Through the evolution and deployment of UCR data, certain categories and types of data have become normative mechanisms for evaluating and understanding crime. Normative data categories are used to delimit not only what is considered necessary knowledge for the public but also what is considered *possible* to know about crime and law enforcement. These ideological boundaries find their material counterparts through federal resource allocation and management, which dictate policy and practice. However, despite the crystallization of normative data categories, the federal centralization of data collection and the supervisory powers of the FBI, policing agencies' submission of data to the UCR remains voluntary and ad hoc. This feature of such federal data collection efforts means that consequences for agencies that either do not report data or report only partial data are minimal. Purposeful and organized data collection requires significant infrastructure and resources. The lack of oversight or enforcement of data collection standards gives law enforcement agencies little incentive to invest in these practices and leaves the historical record around law enforcement with only the impression of being comprehensive and meaningful.

Even as methodologies and practices around the analysis of UCR data have greatly expanded and shifted, some constants have remained: 1) officials, journalists, social scientists and many citizens understand the role of the federal government to include the gathering of data on policing; 2) despite the legal foundations granting power to the Attorney General and the formalization of data collection, the supply of data from individual law enforcement agencies remains voluntary, preventing the uniformity in the corpus of data collected that the UCR's name promises; and 3) the standard categories of 'police data' that get collected tend to exclude data about the activities of law enforcement officials themselves.

COMPSTAT and the Emergence of 'Evidence-Based' Policing

The UCR worked to shape the normative definitions and practices around law enforcement data collection at the federal level, but did little to enforce a relationship between those data collection practices and the everyday labor of law enforcement. In the mid-1990s, an initiative called CompStat (Compare Statistics) emerged that placed data collection and analysis at the heart of police work, tying performance metrics to all aspects of understanding and responding to crime. This program, much like the UCR, both reflected and shaped particular ideological commitments concerning crime, but more than the UCR, it embedded those commitments into the fabric of police labor. Although many of those commitments and assumptions were outgrowths of previous police practice, the use of CompStat engendered significant shifts, including: 1) expanding the mission of policing to encompass crime prevention; 2) prioritizing the expertise and practices of larger, urban law enforcement agencies; 3) increasing emphasis on management; and 4) developing data 'dashboards' that were accessible to the public.

CompStat is now considered shorthand for a broad range of coordinated activities that integrate statistical science into policing practice in a holistic way. While CompStat is not and has never been a single computer program, its antecedent can be traced to a specific project led by Jack Maple of the New York Police Department called Charts of the Future, which attempted to deal with transit-related crime systematically (Henry 2002; Vlahos 2012). The tool was expanded when William J. Bratton—previously the Chief of New York City Transit Police—was appointed in 1994 to be the New York City Police Commissioner. During the two years of his first term as commissioner, the program grew from a rudimentary reporting system to the heavily integrated system and series of practices it now represents. A core component of CompStat became regular meetings, which included personnel from every one of New York's 77 precincts, 12 transit districts, and nine police service areas. Together they compiled a summary of crime data and implemented statistical analysis to identify crime patterns and account for the use of police resources. Geolocated data was given to the chief of the CompStat Unit, who uploaded it into a city-wide database and generated weekly, monthly, and yearly city-wide reports. The weekly meetings that accompanied these reports put unit and precinct commanders together and tied patterns of criminal activity with specific tactics and models of leadership.

The introduction of CompStat into the NYPD ecosystem represented a profound shift in policing culture and practice by establishing a wholesale adoption of and alliance between policing, statistical logic, and an emergent 'police managerial paradigm' (Walsh 2001). A 2013 report published by the Police Executive Research Forum (PERF) and the Bureau of Justice Assistance characterized the penetration of CompStat into their member law enforcement agencies at just under 80%, with 6% reporting that they intended to integrate CompStat into their department in the future (Bureau of Justice Assistance and Police Executive

Research Forum). In the same report, the criminologist George Kelling, who posited the 'broken windows' theory of crime, is quoted as saying, 'Departments don't have to justify doing Compstat. They have to justify not doing Compstat. The gains Compstat has made in policing are obvious.' Broken windows theory stated that what Kelling and James Q. Wilson (1982) identified as visible signs of disorder, such as broken windows, were the precursors to criminality. This theory became the foundation for new tactics for law enforcement focusing on smaller crimes, constant patrolling and stop and frisk practices (Fagan and Davies 2000).

On the one hand, CompStat enabled policing agencies to tie so-called 'quality of life' and performance indicators to the data produced and analyzed. On the other hand, it also enabled police agencies to control how much access the public could have to a set of carefully mapped (but largely decontextualized) data. It also served as a rhetorical device to justify continued or heightened police presence based on location and made it possible to link crime statistics to property values—which, in turned, shaped housing development initiatives and collective narratives about the character of particular neighborhoods or geographical divisions in a given city. CompStat also influenced policing policies, resource allocation decisions, and policing tactics that were informed by the desire for the production of data that could show *improvement*. Among these tactics were the controversial 'stop and frisk' and 'broken windows policing' tactics pioneered by NYPD Commissioner Bratton, which relied on data collection about low-level crimes such as turnstile jumping and panhandling, which are smaller, more visible, easier to count, and easier to report. But by focusing on this kind of criminal activity, NYPD officers had also increased the number of people—especially people of color—with whom they made contact. Thus, CompStat, and the uncritical approach to data collection it relied upon, helped to further entrench racial discrimination in policing, and encouraged a wide range of abuses. Several scholars have documented the ways in which the CompStat system newly incentivized glaring abuses, including the falsification of crime figures in order to elude scrutiny and an increase in the violent tactics used by police in communities of color (Silverman 2012; Schoolcraft 2013).

This overview is not an attempt to characterize abuses of power or racially motivated police action as products of Compstat; rather, its purpose is to locate a shift in rhetoric that justified the continued or increased over-policing of communities of color through the gathering and performance of data. CompStat implementation gave rise to a broader phenomenon of evidence-based policing in the late 1990s and early 2000s, a strategy that fundamentally relied on data gathering and analysis as a core element of police work. According to this strategy, raw material for targeting, testing and tracking phenomena was tied to specific populations, locations and times (Sherman 2013). But the use of police data as evidence and justification for future action assumed the integrity and availability of quality data—and as several scholars have noted, this data was often skewed, incomplete and again entirely focused on crime. Overall, CompStat normalized the performance of data-driven policing focused on crime prevention at the precinct and city level.

Even as Compstat has become integrated into law enforcement agencies across the United States, and consistent data gathering has become an integral part of police work, national data gathering efforts have continued to rely on ad hoc self-reporting to identify national trends in both crime and policing well into the twenty-first century. Although there has been a new reporting standard since 2001, the National Incident Based Reporting System (NIBRS) still relies on self-reporting. NIBRS is meant to zero in on specific data points rather than focusing on the national aggregate. Incidents reported include contextual data such as location, temporal and status data. Adherence to and adoption of NIBRS can be costly, and it also has side effects for agencies that adopt the system, including inflated crime data due to more robust data collection practices. This inflated data can significantly skew the understanding of criminalized behavior in particular communities. The FBI has declared that transitioning states to NIBRS from UCR reporting is a top priority in order to hopefully develop more informed policy and management. Individual states have been slow to get on board, but, alongside the Bureau of Justice Statistics and the National Crime Statistics Exchange, the FBI is nonetheless attempting to make NIBRS the standard for data collection by 2021. Since the system relies on self-reporting, resources that could be spent on oversight or enforcement are instead spent trying to convince law enforcement agencies to update their systems. In a short video produced about NIBRS, the FBI simultaneously capitalizes on nostalgia and heroism by showcasing photos of Bonnie and Clyde while emphasizing the need to transition to more contemporary, technically savvy law enforcement techniques ('NIBRS Overview' 2016). As data elements pop up on the screen, the sound of typing dominates the soundtrack, a contrast to the older modes of communication highlighted in the representations of the birth of the UCR, and a voiceover announces that 'it's not 1930 anymore' ('NIBRS 101' 2016). While UCR would only keep track of one offense per incident (something could either be arson OR burglary but not both), NIBRS allows ten offenses to be logged per incident and includes fifty-two offense classifications. This increased granularity responds directly to an increased focus on the identification of patterns through

maximizing data points endemic in CompStat programs. In 2015, then-Director of the FBI James Comey stated that NIBRS gives law enforcement the ability to recognize and understand patterns and trends rather than counting individual incidents and, in this way, aids crime prevention (Comey).

The UCR currently produces four types of summary data: Offenses Known and Cleared by Arrest, Property Stolen and Recovered, Supplementary Homicide Reports, and Law Enforcement Officers Killed or Assaulted. One of the most remarkable features of the Law Enforcement Officers Killed and Assaulted (LEOKA) report is the number of caveats and warnings accompanying the landing page for each report. In a section entitled 'Data considerations,' specific issues of data type and vocabulary are pointed out in order to facilitate use of the published data. The considerations include pointing out that different methodologies are used for collecting and reporting data about officers who have been assaulted and officers who were killed as well as an explicit call to attention regarding the way in which weapons are defined: 'The UCR Program considers any parts of the body that can be used as weapons (such as hands, fists, or feet) to be personal weapons and designates them as such in its data' (Federal Bureau of Investigation 2017). This level of detail and attention to data concerning violence against police officers contrasts the persistent lack of data around in-custody deaths, use of force and shooting data. Several data types are separated out from the general UCR reporting, including Hate Crime Statistics, Human Trafficking, and Cargo Theft. These differences in methodology, as well as the expansion of terminology regarding weapons, has the potential to contribute to the inflation of statistics concerning violence against officers, particularly with regard to resisting arrest. The focus within the report on explaining data considerations and the attention to detail regarding methodologies enable LEOKA data to be analyzed with a level of nuance not given to other data types.

Data Collection about the Actions of Police

Data collection about the actions of law enforcement officers has a history distinct from that of data about street crime, and it is often both reliant on vague categories and governed by a distinct set of methodologies and policies. For instance, data collected about the actions of law enforcement officers often comes with many layers of methodological caveats. At the same time, public data collection efforts typically occlude the identity of any specific law enforcement officer involved, making the identification of behavioral patterns impossible for anyone outside of an individual agency.

At the federal level, data about police actions is collected by several agencies and in several forms. The Bureau of Justice Statistics (BJS), for instance, has collected data on contact between the police and the public through a variety of surveys and programs since 1996. As of this writing, the latest comprehensive data is available only through 2011 and relies heavily on the National Crime Victimization Survey (NCVS). Rather than relying on reporting from law enforcement agencies (as the UCR does), the NCVS compiles data from a survey of approximately 70,000 households randomly selected from a stratified multistage cluster sample (Bureau of Justice Statistics 2016). The United States Census Bureau then conducts interviews with the selected households about a range of topics related to criminality, and attitudes towards crime and policing.

Collection of law enforcement 'use of force' statistics is, in fact, a mandated responsibility of the United States Attorney General. This mandate comes under the Violent Crime Control and Law Enforcement Act of 1994 (H.R. 3344, Pub.L. 103–322). Details are outlined within Subtitle D: Police Pattern or Practice, Section 210402 (H.R. 3344, Pub.L. 103–322), which charges the Attorney General with the gathering 'through appropriate means' of statistics on the use of excessive force by law enforcement officers. This section also specifically calls attention to the limitations of data use and data elements. Specifically, it notes that '[d]ata acquired under this section shall be used only for research or statistical purposes and may not contain any information that may reveal the identity of the victim or any law enforcement officer' (H.R. 3344, Pub.L. 103–322). As the head of the Department of Justice, the Attorney General's data gathering requirements are met through the efforts of the Bureau of Justice Statistics (which is under the Office of Justice Programs). The BJS offers definitions for 'force,' 'use of force,' and 'excessive force,' but it makes no attempt at documenting the sources for these definitions. Indeed, its distinction between 'use of force' and 'use of excessive force' relies on a common understanding and definition of reason on the part of the officer rather than any particular reference point (Alpert and Smith 1994; Klinger and Brunson 2009). 'Use of force,' for example, is defined as '[t]he amount of effort required by law enforcement to gain compliance from an unwilling subject,' whereas 'the use of excessive force' is defined as '[t]he application of force beyond what is reasonably believed to be necessary to gain compliance from a subject in any given incident' (Bureau of Justice Statistics 2016).

The BJS also periodically attempts to gather a range of law enforcement data that includes salary and demographic information, education and training requirements, computing and information systems, vehicles, special units or initiatives, etc. Done via voluntary questionnaire and representative sampling of state and local law enforcement agencies, the latest report from 2013 for the Law Enforcement Management and Administrative Statistics (LEMAS) represents just under 3,000 law enforcement agencies and relies entirely on self-reporting. Finally, the BJS also operates the Arrest-Related Deaths program (ARD) as an annual national census of persons who died either during the process of arrest or while in custody of law enforcement (Bureau of Justice Statistics 2016). Notably, the terms and definitions used for this data collection separate out the manner of death, classing the death as either homicide or natural. Natural includes '[d]eaths attributed to natural agents such as illness or internal malfunctions of the body. The majority of arrest-related deaths recorded as "natural" were due to heart complications. Other natural deaths included complications from long-term illnesses' (Bureau of Justice Statistics 2016). For contrast, the definition of homicide according to the Bureau of Justice Statistics is 'willful killing of one human being by another' (Bureau of Justice Statistics, 2016). All of these definitions leave little space for any clear observation of the interaction between police and the public, both generally and specifically, and eschew the complexity of the natural processes of the body, ignoring the stress and potential violence of arrest as an aggravating factor to preexisting or dormant physical conditions.

Parallel to the data collection efforts of the BJS, a coordinated data and analysis effort called the Police Data Initiative (PDI) arose. A direct outgrowth of President Obama's Task Force on 21st Century Policing,[2] the PDI was meant to be a public-facing resource rather than an internal data gathering program. The initiative was launched in 2015 as an attempt to create a community of practice around police open data projects and technologies, bringing together law enforcement agencies on a voluntary basis with technologists and researchers with easy-to-use user interface and dashboard style tools. The Police Data Initiative was a challenging project (even among temporarily politically supported projects) because it depended upon the support of the administration for sustainable infrastructure and staff and upon the ability of project members to positively motivate individual law enforcement agencies to participate. Capitalizing on the expertise and motivation of two White House Innovation fellows, the project brought Code for America on board to create accessible and easy-to-use open data portals (Wardell and Ross 2016). After a year, the project boasted participation from 53 jurisdictions. After the 2016 election, the project was moved under the supervision of the Police Foundation, a non-profit whose mission is to 'advance policing through innovation and science' (Police Foundation 2016). The organization was founded in 1970 and has focused on training and funding for police departments making operational or systemic changes. Over the years, the leadership of the Police Foundation has included former police chiefs, people from the private sector and former public officials. The Police Foundation's work is predicated on the idea that policing can be improved through the gathering and analysis of quality data as well as 'innovation.' They identify their mission as the effort to 'empirically and dispassionately examine the issue facing police and to develop, test, and disseminate ideas about how best to deliver police services' (Police Foundation 2016). Their research is presented in a timeline that aligns changes in policy with changing technologies. The foundation consistently produces research on how rank-and-file police officers spend their time, understand and use their resources, and conceptualize their role in society. However, the access afforded to the Police Foundation is due at least partially to the fact that they do not execute any oversight; their research is in service to the practice of policing itself and therefore narrowly conceives its research and potential outcomes. The Board of Directors is made up of a mixture of people from business, law enforcement and the criminal justice system.

Transitioning the Police Data Initiative was necessary for its survival because political support could not be assumed with the change in presidential administration. However, its transition represented a shift from a federally funded and supported project focused on citizen engagement to a program within an organization that understands law enforcement as its sole constituency. The Police Foundation, like the federal government, has a vested interest in the maintenance of the policing system, but the goals and focus of the data gathering and the tools for analysis have changed dramatically. The latest figures suggest that 130 law enforcement agencies are participating in the PDI, making select datasets open and accessible through the Police Foundation's portal (Police Data Initiative 2016). Even with massive support and buy-in from multiple communities, dedicated staff and resources from the White House, 130 law enforcement agencies represent less than ten percent of the estimated 18,000 law enforcement agencies throughout the

[2] As of writing, the full report of the President's Task Force on 21st Century Policing is no longer available on the Department of Justice's Community Oriented Policing Services (COPS) website.

United States (Banks et al. 2016). As yet, there is no specific standard for the datasets volunteered, so it is difficult to quickly compare the depth and breadth of datasets provided by agencies. Some datasets require a login or other verification of identity, all are in different formats and contain different elements, and a few come with disclaimers about comparing these data sources with other, more complete data sources. The interface provided by the Police Data Initiative itself is relatively easy to search and understand, but does not extend to the particular datasets once you follow through.

Finally, in June 2017, the FBI launched the first National Use-of-Force Data Collection Pilot Study (Bruer 2016). The new system includes three types of use-of-force incidents and information related to each incident. The system will include data about the following: when a fatality occurs that is connected to the use of force by a law enforcement officer; when there is serious bodily injury connected to the use of force by a law enforcement officer; and when a firearm is discharged by law enforcement in the direction of a person regardless of whether or not an injury or death occurred (Federal Bureau of Investigation 2017). The system will collect data about the incident, victim and officer. While the system gathers a range of data in order to provide context for the use-of-force incident—including information regarding the circumstance for the initial interaction, how many officers were present, what were the roles of officers involved, etc.—the name or identifying properties of the officer are not captured. This is especially notable because if the use-of-force data is to have impact in the aggregate and address both systemic and individual problems, not capturing the identity of officers makes it that much more difficult to identify and diagnose persistent abusive and violent behavior at a more granular level than the reporting law enforcement agency. As it stands, there is no standard time period for releasing data or reports for use of force by the FBI, and participation remains voluntary for law enforcement agencies. According to the FBI itself, the impetus and organizing around a national effort to collect use-of-force data began in 2015 with a recommendation by the Criminal Justice Information Services Advisory Policy Board (APB) of the FBI and the eventual formation of the National Use-of-Force Collection Task Force in 2016. But the collection of this data could also be a response to coordinated and highly effective calls for increased transparency and accountability precipitated by the organized and collective response of activists to a series of well-publicized deaths at the hands of police. This kind of data, however incomplete, remains outside of a broader framework for redress or citizen empowerment and continues to adhere to the entrenched managerial logics of CompStat.

Citizen Documentation Efforts

Although federal and local agencies have, over the past 100-plus years, developed methods for collecting data about crime, criminal behavior, and related categories, they have rarely mobilized their resources to document, or share available data about, the activities of police officers. But scholars, activists, and others have developed a parallel set of 'citizen data' projects designed to fill in these gaps and to put data existing to work. These efforts go back at least as far as the nineteenth century, when W. E. B. Du Bois began gathering data and creating visualizations for an exhibit he co-curated (with the Assistant Librarian of Congress Daniel Murray and the lawyer Thomas J. Calloway) for the 1900 World's Fair. 'The Exhibit of American Negroes' included registered patents, photographs, a bibliography, information from Historically Black Colleges and Universities (HBCUs), as well as hand-drawn visualizations of data gathered by Du Bois and his students at Atlanta University. The exhibit in general and these visualizations in particular placed information gathering and display as a means of visibility, empowerment and legitimation (Library of Congress n.d.).

A more recent example of citizen-led data collection and visualization work is Campaign Zero. This project arose out of the efforts of a diverse set of activists, community members and policy experts to articulate a clear connection between both access to and ownership of data, data analysis, and data narratives, and the important role that data and data stories play in entrenched power dynamics. Although the organization and its planning team have had official connections to President Barack Obama's Task Force on 21st Century Policing, the project is an independent effort to coordinate research and organizing tactics in order to gather a wide range of information that is not adequately collected by state agencies (Campaign Zero 2015). By providing well-designed data visualization and supporting research, the site provides entry points for a diverse set of community members and empowers them to participate in local governance. Campaign Zero's information visualizations typically combine statistical analysis with expert graphic design elements to communicate large and complex data sets to dramatic and immediate effect. Additionally, these visualizations are often accompanied by talking points and action items, shortening the distance between information and action. The informational ecology of Du Bois's World's Fair project is distinct from that of Campaign Zero, but both offer a critical stance toward data collection, analysis and display from official, state-sanctioned sources when it comes to representing the lived experience of communities of color.

Scholars Morgan Currie, Britt Paris, Irene Pasquetto and Jennifer Pierre have developed a data augmentation project called the Police Officer-Involved Homicide Data (POIH Data) Project, which deploys data 'produced by local activist groups and news organizations with fewer resources than official federal and municipal entities' (Paris and Pierre 2017; Currie, Paris, Pasquetto and Pierre 2016). The creators of A People's Archive of Police Violence in Cleveland have attempted to create a space for Cleveland citizens to share information, create and share testimony in a 'safe and secure' space and collectively express the impact, affect and effect of police violence (Williams and Drake 2017).

Citizen data collection and production exists within a broader historical context that includes civic engagement, open data and open government initiatives, and citizen science. However, citizen data gathering and analysis concerning policing represents a departure from traditional modes of thinking through participatory data projects or transparency initiatives within the government. In this space, organizations seek to both fill in the informational gaps left by institutional bias skewed towards the maintenance and legitimation of policing as well as amend extant data to better reflect lived experience of communities. For example, an organization like CopWatch explicitly fills the information gap around patterns of abuse regarding individual police officers (Huey, Walby, and Doyle 2006; CopWatch n.d.). As mentioned previously, data concerning individual police officers is not provided by official reporting entities, in order to protect officers. This lack of data fosters a lack of accountability and fails to inform the community about officers who abuse their power consistently. CopWatch uses structured data and video evidence to construct a searchable database of individual police officers. Those who participate argue that the community knows its officers, and accountability at the community level is more effective than at the precinct or city level. By tying individual officers to their patterns of behavior, CopWatch attempts to facilitate a community-led accountability structure based on strong local ties and persistent visibility.

Platform Policing and Aggregation

New data gathering techniques and technologies have recently found application through the implementation of cloud-based evidence management systems connected to a range of emergent tools such as police-worn body cameras. Unprecedented changes to law enforcement practice are driven by a convergence of policy, technological development and resource scarcity. In particular, the rise in police-worn body camera programs has presented substantive infrastructural challenges for law enforcement agencies as they contend with massive data storage requirements, expensive interconnected systems, a cache of video evidence too voluminous to analyze with human eyes, and new practices and requirements for law enforcement expertise. Companies such as Axon (formerly Taser, and the leader in the law enforcement technology market) have begun to transform the labor of policing in general and the production and gathering of data specifically through an integrated platform, a system that ties police workflow together via constant and consistent interaction with digital interfaces that produce data and metadata continuously. Policies around new technologies in the policing space are slow to coalesce and policies governing long-term storage and use and reuse of data are virtually non-existent (Wood 2017). Private contractors working with police is not a new phenomenon; many departments use multiple proprietary records management systems (RMS) to manage disparate data sources ranging from computer aided dispatch (CAD) systems used by emergency services to prison and jail records management systems. Systems such as Evidence.com, however, in an attempt to de-silo these informational environments, lock departments in from the moment of data capture to the entire data lifecycle through to prosecution and any archival use or reporting. Evidence.com is the cloud-based evidence management system that Axon developed using Microsoft Azure Government, which adheres to appropriate security standards for public information (Gupta 2014). This platform is a flexible and robust data-rich environment shared by law enforcement, both prosecuting and defense legal teams, and any other appropriate constituents. While audit trails tying data to individuals throughout the system is a feature of the system, it fails to take into account work cultures and work-arounds extant in the system's affordances. Axon expects to become a nose-to-tail prediction system housing the largest store of police data in history, combining developments in machine learning and artificial intelligence in the hopes of accomplishing richer and quicker data analysis (Taser International 2017). These systems deepen the data paradox of police data because, in contrast to the ad hoc, sluggish reporting process followed by the numerous government efforts outlined previously, evidence management systems like Evidence.com could easily report, aggregate and make data available to the public. In fact, it is clear that this data environment is exactly the future source of revenue for Axon and other such companies because the real investment of law enforcement agencies is in the subscription services rather than the devices used in the field or in the office.

This data-rich environment presents us with numerous challenges and potential consequences arising from the large-scale implementation of closed, proprietary systems that produce, maintain and analyze data.

1. Access and maintenance of law enforcement information
   Law enforcement policies governing access to information are idiosyncratic and exist in varying stages of development. While some are regulated by state recordkeeping laws governing access and public disclosure, some are ruled by agency policy and may present a challenge for the public if their policies are ambiguous. Although proprietary systems within law enforcement are not new, there is a significant risk presented by a single company governing all aspects of law enforcement records management. For one, in the absence of significant expertise regarding information management or state law, agreements between local law enforcement and a contractor might overlook vital details of a data management plan. For example, given the cost of data storage and expertise required, consequences might be catastrophic if a law enforcement agency no longer wanted to contract with a particular company, or if that company simply went out of business.
2. The shifting of legal categories
   The integration of a cloud-based records management system that manages information beginning with the first moment a member of the public interacts with a law enforcement officer also challenges pre-existing legal categories. For example, police notebooks, informal documents that are historically the most detailed record of police activity, remain outside of the public purview because they are not subject to subpoena (although this varies from state to state). The record of events, the formal document, has been a report often written after reflection and coordination, subsequently approved by a supervisor (Baer and Armao 1995). However, a cloud-based records management system that structures observations, notes and commentary into a formal and connected system might force a re-thinking of these legal categories. This system compresses the time and distance typically found between an informal note taking system in the field and a formal report writing system that happens after the fact. If an integrated cloud-based system for documentation and records management is constantly available, there is no need for formal transcription and translation from the informal to the formal.
3. Over-representation of large city law enforcement through data
   One could argue that CompStat already represented a move toward privileging tools developed in large-scale urban environments and assuming their utility across agencies of varying size, but the services associated with connected platforms entrench this further, particularly with respect to integrating machine learning. Machine learning requires massive datasets, most easily produced by large urban law enforcement agencies. According to the US Department of Justice, as of 2013, almost half of the country's law enforcement departments employed fewer than 10 officers (Reaves 2015). Smaller departments also have fewer resources, making the out-of-the-box solutions presented by a cloud-based evidence management system highly attractive. In other words, systems optimized based on the needs and practices of larger departments will increasingly become the standard available for all agencies regardless of size and local context.
4. Unprecedented connection and the blurring of public/private barriers
   The past decade has already seen unprecedented cooperation between private companies and public forms of surveillance, breaking down distinctions between commercial and law enforcement data collection (Bellovin et al. 2016; Walden 2013). However, the flow of law enforcement data for commercial research and design stretches current conversations around privacy that privilege the disclosure of identifiable information, and it enshrines user responsibility. These dynamics shift significantly when considering the use of crime data to train artificial intelligence regardless of the identifiability of that data. It can be enormously difficult to remove information from police databases, even when that data is admittedly incorrect (Kanno-Youngs and Porter Jr. 2018). With increasing aggregation and movement of data across commercial sources, incorrect data can have impact across sectors. Additionally, companies currently managing data may be more or less secure and without explicit exit or data migration strategies, long-term integrity of data could be a significant problem.
5. The entrenchment of techno-deterministic logic and rhetoric
   In the overlap between technology and law enforcement, there is a strong techno-determinist impulse to associate a new technology with the solution to a larger systemic problem. We can see this phenomenon when we consider the rhetorical registers within which police-worn body cameras have operated between 2014 and 2018. Their deployment was met with enthusiasm because people saw the cameras as less

biased than individual officers, a relatively neutral observing party. However, in practice, these cameras are enmeshed in complex histories, current policies, pre-existing biases encoded technically, etc., all leading to the conclusion that they are not the solution to excessive force and police violence. Similarly, algorithmic risk assessment, particularly with respect to recidivism, has touted a mechanical objectivity, a distance from bias; yet, as studies have shown, these systems replicate, entrench and embolden pre-existing biases (Angwin 2016; Wood 2016). Automating all aspects of police information gathering, management and analysis has the risk of extending this techno-deterministic logic that posits machine-created information as more objective than human-generated information.

Conclusion

In spite of a century of police data gathering, the category of 'data about policing' can nonetheless be understood as endangered. Early efforts at categorizing and documenting crime at the federal level laid the groundwork for the expectation that 'police data' would be understood exclusively as data about street crimes and would mainly be used internally by policing agencies themselves rather than as a public record collected by state agencies to support accountability and public oversight. Even those data collection programs mandated by federal law still operate without consistent support or empowered oversight, data subject to political winds. Documentation is historically an active element of policing, a purposeful record of an officer upon an encounter with a citizen. Increasingly, the implementation of evidence management systems and policing platforms accelerates problems with documentation and data collection. The contemporary paradox is one of vast data stores creating economic value in a secondary police data market that remain unavailable to the public for scrutiny.

Competing Interests

The author has no competing interests to declare.

References

Alpert, Geoffrey P., and William C. Smith. 1994. "How Reasonable Is the Reasonable Man: Police and Excessive Force." *Journal of Criminal Law and Criminology*, 85(2): 481–501. Autumn. DOI: https://doi.org/10.2307/1144107

Angwin, Julia. 2016. "Make Algorithms Accountable." *New York Times*, August 1, 2016. https://www.nytimes.com/2016/08/01/opinion/make-algorithms-accountable.html.

Arcidiacono, Davide, and Giuseppe Reale. 2017. "The Open Data Movement: Young Activists between Data Disclosure and Digital Reputation." *Partecipazione e Conflitto*, 9(3): 918–48. https://siba-ese.unisalento.it/index.php/paco/article/view/16661/14324.

Axon. n.d. "Evidence.com." Accessed January 10, 2018. https://www.axon.com/products/evidence.

Baer, Harold, Jr., and Joseph P. Armao. 1995. "The Mollen Commission Report: An Overview." *New York Law School Law Review*, 40(1–2): 73.

Banks, Duren, Joshua Hendrix, Matthew Hickman, and Tracey Kyckelhahn. 2016. "Program Report." *National Sources of Law Enforcement Employment Data*. NCJ 249681, Washington, DC: US Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. https://www.researchconnections.org/icpsrweb/instructors/biblio/series/169/resources/121823?sortBy=1.

Barnes, Harry Elmer. 1921. "The Historical Origin of the Prison System in America."*Journal of the American Institute of Criminal Law and Criminology*, 12(1): 35–60. May. DOI: https://doi.org/10.2307/1133652

Bellovin, Steven M., Matt Blaze, Susan Landau, and Stephanie K. Pell. 2016. "It's Too Complicated: The Technological Implications of IP-based Communications on Content/Non-content Distinctions and the Third-Party Doctrine."*Harvard Journal of Law and Technology*, 30(1). September.

Bruer, Wesley. 2016. "DOJ to enable nationwide collection of use-of-force data." *CNN Politics*, October 13, 2016. https://www.cnn.com/2016/10/13/politics/police-force-justice-department-loretta-lynch/index.html.

Campaign Zero. 2015. "We Can End Police Violence in America." https://www.joincampaignzero.org/vision/.

Comey, James. 2015. "Speech at the International Association of Chiefs of Police." Published October 28, 2015 by The IACP. Accessed February 2, 2018. https://www.youtube.com/watch?v=wxNnCpOX90M.

COPS Office. 2015. "President's Task Force on 21st Century Policing Implementation Guide: Moving from Recommendations to Action." Washington, DC: Office of Community Oriented Policing Services.

Copwatch. n.d. Accessed January 10, 2018. http://www.copwatch.org/.

Cranefield, Jocelyn, Oliver Robertson, and Gillian Oliver. 2014. "Value in the Mash: Exploring the Benefits, Barriers and Enablers of Open Data Apps." *Proceedings of the European Conference on Information Systems (ECIS)*. Tel Aviv, Israel, June 9–11, 2014. https://aisel.aisnet.org/ecis2014/.

Crunden, Robert Morse. 1982. *Ministers of Reform: The Progressives' Achievement in American Civilization, 1889–1920*. Champaign: University of Illinois Press.

Currie, Morgan, Britt S. Paris, Irene Pasquetto, and Jennifer Pierre. 2016. "The Conundrum of Police Officer-Involved Homicides: Counter-Data in Los Angeles County." *Big Data & Society*, 3(2): 1–14. December. DOI: https://doi.org/10.1177/2053951716663566

Davies, Tim. 2010. "Open Data, Democracy and Public Sector Reform." *A Look at Open Government Data Use from data.gov.uk.* https://www.opendataimpacts.net/report/.

Fagan, Jeffrey, and Garth Davies. 2000. "Street Stops and Broken Windows: Terry, Race, and Disorder in New York City." *Urban Law Journal*, 28(2).

Federal Bureau of Investigation. 2017. "Preliminary Semiannual Uniform Crime Report, January–June, 2017." Accessed January 14, 2018. https://ucr.fbi.gov/crime-in-the-u.s/2017/preliminary-report.

Federal Bureau of Investigation. 2017. "The National Use-of-Force Data Collection." *CJIS Link*, June 15, 2017. Accessed January 17, 2018. https://www.fbi.gov/services/cjis/cjis-link/the-national-use-of-force-data-collection.

Gupta, Shashi Kant, and Faisal Siddiqui. 2014. "Secure Data Storage and Retrieval in the Cloud." *International Journal of Research in Engineering and Technology*, 3(10). https://ijret.org/volumes/2014v03/i22/IJRET20140322008.pdf.

Henry, Vincent E. 2002. *The COMPSTAT Paradigm: Management Accountability in Policing, Business, and the Public Sector*. Flushing, NY: Looseleaf Law Publications.

Huey, Laura, Kevin Walby, and Aaron Doyle. 2006. "Cop Watching in the Downtown Eastside: Exploring the Use of (Counter) Surveillance as a Tool of Resistance." In: *Surveillance and Security: Technological Politics and Power in Everyday Life*, Torin Monahan (ed.), 149–65. New York, NY and London: Routledge.

Jenkins, Philip. 1984. "Eugenics, Crime and Ideology: The Case of Progressive Pennsylvania." *Pennsylvania History: A Journal of Mid-Atlantic Studies*, 51(1): 64–78. January

Johnson, Jeffrey Alan. 2014. "From Open Data to Information Justice." *Ethics and Information Technology*, 16(4): 263–74. DOI: https://doi.org/10.1007/s10676-014-9351-8

Kanno-Youngs, Zolan, and Gerald Porter, Jr. 2018. "Gang Databases are the Latest Controversial Policing Tool." *Wall Street Journal*, July 26, 2018. https://www.wsj.com/articles/gang-databases-are-the-latest-controversial-policing-tool-1532597400.

Kelling, George L., and James Q. Wilson. 1982. "Broken Windows: The Police and Neighborhood Safety." *Atlantic*, March 1982. https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/.

Klinger, David A., and Rod K. Brunson. 2009. "Police Officers' Perceptual Distortions during Lethal Force Situations: Informing the Reasonableness Standard." *Criminology & Public Policy*, 8(1): 117–40. February. DOI: https://doi.org/10.1111/j.1745-9133.2009.00537.x

Library of Congress. "African American Photographs Assembled for 1900 Paris Exposition." Accessed December 12, 2017. https://www.loc.gov/pictures/collection/anedub/.

Meskell, Matthew W. 1999. "An American Resolution: The History of Prisons in the United States from 1777 to 1877." *Stanford Law Review*, 51(4): 839–65. April. DOI: https://doi.org/10.2307/1229442

Milan, Stefania, and Lonneke van der Velden. 2016. "The Alternative Epistemologies of Data Activism." *Digital Culture & Society*, 2(2): 57–74. DOI: https://doi.org/10.14361/dcs-2016-0205

Paris, Britt S., and Jennifer Pierre. 2017. "Naming Experience: Registering Resistance and Mobilizing Change with Qualitative Tools." *InterActions: UCLA Journal of Education and Information Studies*,13(1). https://escholarship.org/uc/item/02d9w4qd.

Poggio, Eugene, Stephen D. Kennedy, Jan M. Chaiken, and Kenneth E. Carlson. 1985. *Blueprint for the Future of the Uniform Crime Reporting Program: Final Report of the UCR Study*. Washington, DC: US Department of Justice, Bureau of Justice Statistics, Federal Bureau of Investigation.

Police Foundation. 2016. "History." Accessed December 14, 2017. https://www.policefoundation.org/.

Reaves, Brian. 2015. "Local Police Departments, 2013: Personnel, Policies, and Practices. Office of Justice Programs Bulletin." Accessed December 10, 2017. https://www.bjs.gov/content/pub/pdf/lpd13ppp.pdf.

Rosen, Lawrence. 1995. "The Creation of the Uniform Crime Report: The Role of Social Science." *Social Science History*, 19(2): 215–38. Summer. DOI: https://doi.org/10.2307/1171511

Schoolcraft, Adrian. 2013. "Right to Remain Silent: Act Two; Is That a Tape Recorder in Your Pocket, or Are You Just Unhappy to See Me?" Interview by Ira Glass. *This American Life*. September 28, 2013. Audio, 41:00. https://www.thisamericanlife.org/414/right-to-remain-silent/act-two-0.

Sherman, Lawrence W. 2013. "The Rise of Evidence-Based Policing: Targeting, Testing, and Tracking." *Crime and Justice*, 42(1): 377–451. DOI: https://doi.org/10.1086/670819

Smith, Megan, and Roy L. Austin, Jr. 2015. "Launching the Police Data Initiative." *White House* (blog), May 15, 2015. https://obamawhitehouse.archives.gov/blog/2015/05/18/launching-police-data-initiative.

Stoler, Ann L. 1989. "Making Empire Respectable: The Politics of Race and Sexual Morality in 20th-Century Colonial Cultures." *American Ethnologist*, 16(4): 634–60. November. DOI: https://doi.org/10.1525/ae.1989.16.4.02a00030

Taser International Inc. 2017. "TASER Makes Two Acquisitions to Create 'Axon AI.'" *PR Newswire*, February 9, 2017. https://www.prnewswire.com/news-releases/taser-makes-two-acquisitions-to-create-axon-ai-300404780.html.

Vlahos, James. 2012. "The Department of Pre-Crime." *Scientific American*, 306(1): 62–67. January. http://www.jstor.org/stable/26014159.

Walden, Ian. 2013. "Accessing Data in the Cloud: The Long Arm of the Law Enforcement Agent." In: *Privacy and Security for Cloud Computing*, Siani Pearson, and George Yee (eds.), 45–71. London: Springer Nature. SpringerLink. DOI: https://doi.org/10.1007/978-1-4471-4189-1_2

Walsh, William F. 2001. "Compstat: An Analysis of an Emerging Police Managerial Paradigm." *Policing: An International Journal of Police Strategies & Management*, 24(3): 347–62. DOI: https://doi.org/10.1108/13639510110401717

Wardell, Clarence, and Denice Ross. 2016. "The Police Data Initiative Year of Progress: How We're Building on the President's Call to Leverage Open Data to Increase Trust between Police and Citizens." *Medium*. https://medium.com/@ObamaWhiteHouse/the-police-data-initiative-year-of-progress-how-we-re-building-on-the-president-s-call-to-leverage-3ac86053e1a9.

Williams, Stacie, and Jarrett Drake. 2017. "Power to the People: Documenting Police Violence in Cleveland." In: "Critical Archival Studies." Michelle Caswell, Ricardo Punzalan, and T-Kay Sangwand (eds.), Special issue, *Journal of Critical Library and Information Studies*, 1(2). DOI: https://doi.org/10.24242/jclis.v1i2.33

Wines, E. C. (ed.) 1871. *Transactions of the National Congress on Penitentiary and Reformatory Discipline Held at Cincinnati, Ohio, October 12–18, 1870*. Albany: Argus Company.

Wood, Stacy. 2016. "Police Body Cameras: Emotional Mediation and the Economies of Visuality." In: *Emotions, Technology, and Design*, Sharon Y. Tettegah, and Safiya Umoja Noble (eds.), 227–39. Cambridge, MA: Academic Press. DOI: https://doi.org/10.1016/B978-0-12-801872-9.00011-9.

Wood, Stacy E. 2017. "Police Body Cameras and Professional Responsibility: Public Records and Private Evidence." *Preservation, Digital Technology & Culture*, 46(1): 41–51. April. DOI: https://doi.org/10.1515/pdtc-2016-0030.

How to cite this article: Wood, Stacy. 2018. The Paradox of Police Data. *KULA: knowledge creation, dissemination, and preservation studies* 2(1): 9. DOI: https://doi.org/10.5334/kula.34

Submitted: 13 March 2018 **Accepted:** 13 October 2018 **Published:** 29 November 2018

Copyright: © 2018 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

]u[*KULA: knowledge creation, dissemination, and preservation studies* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS