Beware of "Big Data Hubris" When It Comes to Police Reform

Christopher Moraff

March 7, 2016

For the past several years police departments around the United States have been betting on "big data" to revolutionize the way they predict, measure and, ideally, prevent crime. Some data scientists are now turning the lens on law enforcement itself in an effort to increase public insight into how well police officers are doing their jobs.

Last year, the city of Indianapolis and Code for America teamed up to launch Project Comport — an open-data platform for sharing information on complaints and use of force incidents. (Nick Selby, a police officer and software developer who consults on policing technology, recently took the system for a ride and wrote about its potential.) And two media projects recently funded by the Knight Foundation focus exclusively on American law enforcement.

The Chicago-based "Citizens Police Data Project" — an initiative of the Invisible Institute — launched a database in November containing more than 56,000 Chicago police misconduct complaints involving thousands of officers. It plans to use its Knight grant to develop a web application to simplify the filing and tracking of complaints. Meanwhile, a project called "Law Order and Algorithms" based at Stanford University plans to collect, analyze and release data on more than 100 million highway patrol stops over the next two years, creating a massive storehouse of police-citizen interactions for journalists and policymakers.

The efforts are part of a larger push for data transparency that has accelerated since controversial police-involved fatalities in Ferguson and Baltimore sparked a national dialogue on police reform.

The nonprofit RAND Corporation has been a strong proponent of data-driven solutions to public safety issues. In late 2014, Nelson Lim, a senior scientist at RAND, called on the Department of Justice to "support police agencies by developing standard protocols for data collection, data sharing and analytical tools police agencies can use" as a means of rebuilding trust in law enforcement.

An open data policy was included among the more than 60 recommendations released last year by the President's Task Force on 21st Century Policing. And in April — at a special

meeting of technologists and law enforcement leaders — the Obama administration launched the Police Data Initiative (PDI) to encourage police departments to provide more of their files for public scrutiny. At last count more than two dozen law enforcement agencies had joined the effort. Collectively, these agencies have released 40 open data sets to date, all of which can be found on the recently launched Public Safety Open Data Portal.

Many of these initiatives are powered by technology from a firm called Socrata, which released a public safety component to its cloud-based open government platform last month and already has contracts with a number of early participants in the Police Data Initiative.

"In the past year we have witnessed a complete shift with many police departments today embracing data transparency as the foundation to enhancing, or in some cases restoring trust," says Kevin Merritt, the company's CEO and founder.

"Complete shift" may be overoptimistic. While the Fraternal Order of Police has expressed tacit support for the Police Data Initiative, some police union officials (most notably the head of the Dallas FOP) have balked at the idea of making civilian complaints publicly available. They mostly cite concerns over privacy and officer safety.

Experts say privacy risks can be mitigated with proper encryption, but getting more departments on board is sure to face resistance.

The Invisible Institute's Jamie Kalven had to sue the city of Chicago to acquire the records that are now part of the Citizens Police Data Project.

Kalven says his goal is to create a national model for "operational transparency" and present a rhetorical challenge to police departments that insist they require secrecy to function.

"What's really significant here is the publicness of the information," he says. "What that makes possible is a dialogue on policing reform that is evidence-based instead of cloaked in secrecy."

That such a need exists is not in dispute. During a speech last year at Georgetown University, FBI Director James B. Comey lamented the lack of data on even something as fundamental as how many people are killed by police officers in a given year.

"The first step to understanding what is really going on in our communities and in our country is to gather more and better data related to those we arrest, those we confront for breaking the law and jeopardizing public safety, and those who confront us," he said.

While the DOJ periodically releases data on police interactions with the public, the two most comprehensive resources on police misconduct — the Cato Institute's policemisconduct.net (curated by Jonathan Blanks), and the work of Bowling Green University's Philip M. Stinson — rely entirely on published media reports.

Ravi Shroff, a research scientist at New York University's Center for Urban Science and Progress and one of the minds behind Law Order and Algorithms, says his team has already collected data on more than 40 million traffic stops with little opposition. He says that while public records requests can be a useful tool for obtaining data, an increasing number of police departments are releasing data of their own volition.

"My opinion is that once stakeholders see the value that results from collecting and releasing more data, this will provide additional impetus [for departments] to create publicly accessible databases," says Shroff.

Unfortunately, with no overarching authority tasked with their oversight, the nation's roughly 18,000 independent law enforcement agencies are pretty much on their own when it comes to collecting and warehousing data. And, exactly what information can be made public is often constrained by individual state privacy laws.

For open data efforts to be successful at a national level, they will require some level of uniformity. The Police Open Data Census, which summarizes open data initiatives in the police departments that have them, shows wide variation among departments on what's included in publicly accessible datasets. The group offers a set of minimum standards for police departments, including providing incident-level data on not only use-of-force, but also civilian complaints, police response times, and information on traffic and pedestrian stops. This data should be up to date and presented in "machine-readable" formats from which data can be extracted.

Success will also mean embracing realistic goals. Some researchers working on the federally sponsored Police Data Initiative envision a massive analytical machine that crunches numbers on everything from an officer's disciplinary record to the number of times he calls out sick, with the goal of proactively weeding out bad apples.

But Kalven warns against putting too much stock in data's ability to facilitate accountability.

"To me there is kind of a data hubris that accepts the notion there is some kind of algorithm that's going to advance police reform," he says. "What we are dealing with is institutions that are fundamentally opposed to addressing the problem. Cities could build just as good a system using index cards if they have someone with the political will to connect the dots."