

This document was produced based on notes taken during the Open Data, the Criminal Justice System, and the Police Data Initiative workshop of the Data & Civil Rights conference. This document represents a general summary of the discussion that took place. Not all attendees were involved in every part of the conversation, nor does this document necessarily reflect the views and beliefs of individual attendees. All workshop participants received workshop materials prior to the event to spark discussion. The primer can be found at: http://www.datacivilrights.org/pubs/2015-1027/Open_Data_Police_Data_Initiative.pdf

Workshop Discussion Notes: Open Data, the Criminal Justice System, and the Police Data Initiative

Overview

This workshop focused on the opportunities and limitations of open data as a driver of transparency and accountability within the criminal justice system. It brought together practitioners, technologies, as well as nonprofits and civil rights groups involved in the collection and analysis of criminal justice data, and/or in the mobilization of community groups to hold police accountable and spur criminal justice reform. Participants sought to evaluate early efforts to open criminal justice data, and to discuss potential opportunities and challenges of using this method as a way to increase transparency into police actions. Prior to a discussion, participants were asked to engage in a “spectrum” exercise to establish where they fell on issues of open data in police accountability. Though there was a general consensus on the benefit of having more information about police actions available to the public, there was significant disagreement about what, how, and by whom data should be collected and released.

Workshop Part 1: Identifying Opportunities and Challenges for Open Data in Policing

At the beginning of the session, participants were asked to identify the opportunities and challenges of opening police data. A number of themes began to emerge. Participants repeatedly pointed to using open data to increase citizen and community engagement. While some saw a potential benefit for citizens to hold police accountable, others thought increased information flows may prompt citizens to be more sympathetic to the challenges of policing work.

Several participants prioritized the need to identify potential datasets to be opened, as well as to improve data that has already been released and make open data more usable. This has been the goal of efforts like

the Police Data Initiative – the White House’s initiative to open up police data on police-community interactions. These participants focused on the need to develop real-time keys to calls for service data, and the need for a set of uniform standards for the data collection process.

The majority of participants saw increased transparency and accountability as the main opportunity that could potentially be provided by open data. As part of this, participants saw open data as playing an essential role as an interface that could be used to improve community relations, to help community groups hold police accountable for actions committed on citizens, as well as to identify police officers who would be more likely to commit unjustified acts of violence. Related discussions explored whether open data could potentially help reduce racially biased policing, and lead to higher trust of police among discriminated populations. Open data was proposed as a way to “change the conversation” about police practices, where released data could be seen as an opportunity to move from “anecdote” to “data-driven conversations” to assess police actions.

Though participants expressed great hope for open data, they also acknowledged challenges and barriers. The majority of concerns centered on existing data availability and quality issues plaguing the criminal justice system, especially on issues related to police and citizen interactions, such as use of force. Issues like data inaccuracies and gaps, as well as inconsistencies in collection and reporting of data across jurisdictions, was used to draw attention to the amount of work that would have to be done to ensure the validity or reliability of any data released as part of open data programs. Some participants raised concerns with data quality, saying that in many cases it is difficult to ensure accuracy in data collection methods, and that in most cases, data about issues like use of force are of poor quality or non-existent.

Alongside these concerns, participants highlighted several potential “gatekeeping” challenges, focusing on how the goals and aims of police departments and unions collecting and releasing open data may conflict with the aims of organizations using the data to hold police accountable. The issue of conscious or unconscious manipulation of data also drew attention to potential biases that could have informed data collection efforts in the past. This data could be harmful without a clear understanding of how bias enters into the data collection process. Privacy issues, such as the potential targeting of officers and victims through the use of open data, highlighted a need for more research into anonymization, and the rights of citizens and officers in open criminal justice data programs.

The high barrier to using open data, which requires a set of technical skills and education and, in some cases, access to increased computer processing power, was also identified as a potential challenge that could prevent or deter more heavily policed communities from using open data to build their own narratives of police and citizen interactions. This last point began a discussion about whether open data and other reforms to increase transparency could solve the underlying causes of police violence, such as systemic and pervasive racism.

The Open Data Spectrum

The majority of the first workshop session was spent on a spectrum exercise, wherein participants identified whether they agreed or disagreed with a set of questions and statements asked by the two workshop moderators. The purpose of the exercise was to expose the group to the diversity of viewpoints present in the room, and to use these areas of agreement or disagreement to foster discussion and debate.

1) Agree or Disagree: Open Data is Dangerous

Those who agreed strongly with the statement pushed back on claims that criminal justice data could be “neutral” and focused on biases at the stage of data collection. They argued that since minority and low-income populations are more observed by police than other populations, releasing data and claiming it is “objective” could be misleading and potentially harmful if used to build larger policy decisions that would reproduce biases present in the initial policing and data collection practices. One participant pointed to socioeconomic biases in 311 data as an example, saying that, if one was to analyze existing pothole data, they would draw the conclusion that potholes only existed in higher-income neighborhoods, as these individuals are the most likely to use 311 services.

Participants who strongly disagreed with the statement reiterated their belief in the power of openness as a driver of transparency, arguing that increasing and enforcing standards for data collection could help with claims that released data was biased or manipulated. One individual who disagreed with the statement for other reasons said that opening data, despite potential quality concerns, would highlight the *lack* of data available, and spur efforts to increase data collection in the criminal justice system.

Those who neither agreed nor disagreed wholly with the statement wondered whether the data that was being released to the public was sufficiently reflective of police and community interactions. One individual said that knowledge from community members such as high school students about their perception of police, might be more indicative of community tensions and areas for intervention than knowing who is arrested by police. Others focused on privacy concerns they would like to be addressed before allowing widespread releases of data.

2) Agree or Disagree: Open Data initiatives are the worst way to improve community-police relations

Almost all participants disagreed with this statement, arguing that there are a number of other initiatives that would be more detrimental to community-police relations. Of those who only moderately disagreed with the statement, one said that, while open data can enable a dialogue, it will only play a small part to build trust between communities and their police.

3) Agree or Disagree: We have good models/examples of how to use open data for police accountability

Attendees who agreed with the statement pointed to the insights gained through the release of NYPD stop-and-frisk data, which had been integral to show a racial bias in policing in New York City. Others looked to transparency efforts that had been undertaken by police in Los Angeles and Los Angeles County. Those who disagreed argued that while data had increased information, it did not lead to significant interventions. In this case, though there have been good models that demonstrate how open data can drive dialogue and debate, there is little evidence that its release led to changing policies and practices. One person who neither agreed nor disagreed fully with the statement said that there is a need to be realistic with the end goal of open data, and it cannot be heralded as the be-all-and-end-all solution.

4) Agree or Disagree: The tradeoff between transparency and privacy is a myth

A majority of the participants disagreed with the statement, with one saying it is “absolutely a trade-off.” One said holding police accountable presents a particular privacy issue in that police can accuse innocent individuals of crime, who may then be included within open datasets, and if de-anonymized, could have potentially negative social consequences or collateral consequences. Without including this type of police-community interaction data, however, it would be difficult to gain an accurate picture about bias in policing. Others highlighted the role of the federal government in providing guidelines (including database models) as to how to ensure anonymity, over local police departments who might not have the technical capacity or know-how to ensure that privacy standards are maintained and enforced.

Those that were in the middle of the spectrum thought that focusing on this tradeoff occasionally detracted from other issues. For instance, one individual said that institutions sometimes use “privacy concerns” as a cover to only release data that reflects well on the institution.

5) Agree or Disagree: Data collection and release is primarily the responsibility of police departments

Those who answered in the affirmative said that the responsibility to collect and release data begins with the police. Those who disagreed offered up alternative solutions, such as a third party who would be responsible for releasing data, or to empower community organizations and independent accountability mechanisms or monitors who can provide oversight and audit the data collection process and what is released. One individual said that while the police should be responsible for releasing data, the decision over *what data should be released* should not be up to individual police departments to decide.

Workshop Part 2: Expanding on Opportunities and Challenges

Issues with Data Quality and Consistency – The Role of Independent Agencies and Organizations

One of the bigger challenges with open criminal justice data is that data on police-community interactions is not being collected or reported consistently. Data that is collected is often incomplete or of poor quality. One participant spoke to their experience assessing quality, bias and gaps in data collection by police. As their organization discovered, the data provided by police departments requires significant cleaning to be usable. Organizational challenges that will enable third parties, such as this participant's institution, to conduct "climate assessment" of data collection and organization, present significant barriers.

Ensuring variables are being consistently defined and measured across jurisdictions also presents a challenge. This means that terms like "use of force" need to be harmonized and standardized across police departments. For example, auditing processes need to be put in place to ensure data is being collected accurately and without bias, including having multiple people look at the data and the creation of reports from data to ensure it is measuring what it purports to measure. Even with these safeguards, the participant said, it is difficult to eradicate issues like selection bias creating more systemic issues in long-term data collection. However, the participant said that outside organizations should play a role in simplifying the process of data collection, transfer, cleaning, and analysis.

Other outside organizations collecting data on police actions, by using media sources instead of data provided or collected from police departments, wondered how their role would evolve without access to more data. Though finding data from media on issues such as deaths by police was possible, obtaining other data independently, such as on traffic stops or stop-and-frisk, is not possible without cooperation from PDs.

Assessing 'Openness' – Increasing Access to Open Data

Other participants also stressed the need to gain input from community members in processes of data collection and ensuring quality and consistency. One challenge that was noted repeatedly in the workshop was open data's high barrier to access, which often requires technical and educational knowledge to assess and analyze which may not be possessed by a majority of citizens.

A number of participants discussed the need to engage in skill-building in communities, or to build better data interfaces, to ensure that community members would use open data to keep police departments accountable. Another participant who had been working with criminal justice organizations to increase access to data on traffic stops, found that 'access' was a relative term, and that community groups could not use the data. Determining the operational definition of 'open' in relation to access and use was an important point made throughout the discussions. One individual said that, in addition to technical and

legal issues affecting openness, access is also a matter of *engagement*. If open data isn't being brought to community members, particularly those that are affected by issues with policing, most won't know the data exists.

To that end, participants representing community and citizen advocacy groups shared their experiences of working with police to improve information flows. One discussed how regular meetings with police commissions on crimes occurring in their neighborhoods, as well as timely and periodic metrics of misconduct complaints, as well as the provision of data around dispositions of complaints, has served to generate engagement from citizens concerned about the goings-on in their communities and support community members' allegations of policing in their neighborhoods with complementary data.

Interoperability across jurisdictions and agencies

In addition to a widespread consensus that data collected on police-community interactions needs to be more robust and of a higher quality, workshop discussions also centered on the need to promote interoperability between jurisdictions, as well as agencies handling different aspects of the criminal justice process. One major concern that was expressed was that data points are frequently not connected – even for the same individual within the system. When looking at arrest data, it is very difficult to know the why the event was initiated, the outcome of the arrest, and whether it was thrown out in court. Tracing data from arrest to conviction (or lack thereof) may help shed light on areas where reform needs to be targeted.

Participants who had experience with data collection on police actions – either through crowdsourced efforts, or through speaking to affected communities – discussed their methods to ensure validity and reliability of data. For the most part, triangulating data through relying on multiple sources (such as media, complaints and grievance reports), combined with surveying affected populations, can help lead to more robust data collection in these areas.

There were ongoing discussions around who determines what types of data is relevant, how it is gathered, and why that data is useful. Additionally, there was still a lack of consensus over what methods should be used to define terms that form the basis for data collection efforts. Part of this issue arose from the different types of needs expressed by various community members, ranging from police oversight professionals, accountability experts, advocates, among other, and what is needed from the data across different audiences. Moreover, even for data sets that are seemingly narrow, there can be discrepancies in understandings of what should be included (i.e. there may even exist variables with regards to how you define and collect data on issues such as 'police violence' and even 'police killings,' and as such, make comparative analysis difficult. There may even exist different desires for types of data, i.e. information around systemic issues (that is, variables that could affect crime and policing – such as access to playgrounds, education, healthcare and other social services) may be desired instead of just use of force data. To that end, participants asked whether the data considered useful for police departments is also the same type of data that would be useful for communities. Questions centered on whether we were collecting

the right metrics, how we come to a consensus on what those metrics are, and who gets to control the process of data collection as programs progress.

Regardless, by the end of the workshop, most participants agreed that, to know how to make black youths feel safest from police, arrest data is not enough. Rather, instead of relying on data coming from the police, we need more data sources which could indicate perspectives and experiences from communities affected by policing practices – including how they feel during and following interactions (i.e. their perceptions of interacting with police), collateral consequence and outcomes of policing encounters, their evaluations of performances of members of service, and their feelings of dissatisfaction about how policing takes place in their neighborhoods.

Some participants offered other metrics to gauge police-community interactions instead, such as performance management metrics coming out of police departments, as well as from surveys of community members.