NNEDV
NATIONAL NETWORK TO END DOMESTIC VIOLENCE

Open Police Data Initiatives:
What Domestic Violence & Sexual Assault Programs Need to Know

In the wake of several high-profile cases of officer-involved shootings and a growing public distrust in law enforcement agencies, the public is demanding accountability and transparency. In 2015, the President's Task Force on 21st Century Policing released a report recommending that law enforcement agencies embrace a culture of transparency by making police data available online on stops, summons, arrests, reported crime, and other law enforcement aggregated data.

One of the proposed initiatives to encourage communication and transparency between law enforcement and communities is the White House Police Data Initiative (PDI), an organized effort to make consistent law enforcement data available to the public. Based on recommendations from the President's Task Force on 21st Century Policing report, PDI is working with law enforcement agencies to publish police data online to provide the public with a deeper understanding of their local law enforcement agency's practices, policies, and day-to-day operations.

In addition to PDI, many law enforcement agencies across the United States have been and continue to release police data online. Agencies are publishing police data in an effort to meet increased requests for public records and as part of city-wide open government initiatives. Police data that are being released include use of force, police complaints, 911 calls, community meetings, trainings, arrest reports, officer-involved shootings, body-worn camera and dash camera use, and crime incidents.

Benefits

For the domestic violence and sexual assault community, police data could help inform the public how law enforcement responds to and handles sexual assault and domestic violence crimes. For example, there has been very interesting analysis of using open police data to reveal law enforcement agency's handling of rape kits and law enforcement investigated sexual assault cases. There may be other potential uses of open police data that could be illuminating in regards to sexual assault and domestic violence.

Concerns

While there may be benefits in using open data to analyze law enforcement's responses to domestic violence and sexual assault, the data sets themselves can be problematic for victims of violence. For example, most of the data sets are incident level, which means that what is published is a specific crime that includes, at the very least, the date/time, location (actual location or block address), incident number, and type of crime (domestic violence, assault, rape victim under age 12, etc.). Since each jurisdiction handles their data differently, some jurisdictions include more data elements such as gender and race of perpetrator, name of responding officer, and a brief description of the crime. This level of preciseness makes it likely that a victim's identity could be revealed.

This project was supported by Grant No. 2013-TA-AX-K006 awarded by the Office on Violence Against Women, U.S. Department of Justice (USDOJ OVW). The opinions, findings, conclusions, and recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the USDOJ OVW. If you have questions, please email: safetynet@nnedv.org

Another concern is that these data sets are published online, making it available to anyone. This could include the perpetrator, nosy friends and family, or employers and landlords. For survivors with high privacy risks, including those who may be in hiding from an abusive offender, they may decide to not call 911 or seek help from law enforcement for fear that that interaction will be published online.

What Can Domestic Violence/Sexual Assault Programs and Advocates Do?

Publishing police data online and PDI could be helpful to the public, researchers, and domestic violence and sexual assault programs in evaluating the how law enforcement approaches domestic violence and sexual assault cases. However, to ensure that it is done safely for victims and takes into account privacy concerns, domestic violence and sexual assault advocates are essential. They can help:

- Identify challenges that victims face when they engage with law enforcement and how data could support or explain those challenges.
- Identify system failures for victims of violence and how law enforcement data could identify reasons for those failures.
- Ensure that data sets do not compromise victim privacy and safety.

What Can You Do?

Know if and how your local law enforcement agency is publishing police data online. Because each jurisdiction is releasing different types of data, look to see what your local law enforcement agency is publishing. To learn if your local law enforcement agency is participating in the White House's PDI, visit this webiste: https://publicsafetydataportal.org/. Another way to see if your local agencies are publishing data online is to google: [your city] + [crime data].

Educate others about why survivor privacy is important. Help law enforcement and community members understand how privacy can impact a victim's ability to maintain their safety. The dynamics of domestic violence and sexual assault are complex, and many may not understand how lack of privacy could deter victims from reaching out for help.

Know what privacy protections exist for victims of domestic violence and sexual assault. Learn about the open record laws in your state. In most states, general police data are considered public record, and in some instances the police data sets that law enforcement agencies are releasing are considered public record. However, publishing potentially identifying police data pre-emptively to the entire world could generate unintended consequences, particularly if the data includes information about domestic violence and sexual assault victims. Look to see if your state laws have exceptions or protections for victims of domestic violence and sexual assault and advocate for their privacy. Contact your state coalition against domestic violence or state coalition against sexual assault for more information on privacy laws specific to victims of domestic violence and sexual assault.

Be a part of the conversation. If the data could possibly reveal a victim's identity or be harmful to a victim's safety, reach out to your local law enforcement PDI team and see if they can adjust their data

This project was supported by Grant No. 2013-TA-AX-K006 awarded by the Office on Violence Against Women, U.S. Department of Justice (USDOJ OVW). The opinions, findings, conclusions, and recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the USDOJ OVW. If you have questions, please email: safetynet@nnedv.org

release to protect victim privacy. Work with your local law enforcement agency and other community members to determine how PDI data might be helpful for survivors. Agencies that are part of the White House PDI are committed to working with their communities to ensure that the data they release is helpful and to discuss issues in their community. Below are several suggestions that could assist in these conversations with the goal of developing policies and practices that support victim privacy.

Minimize Re-identification Risks in Incident-Level Data

In general, it is preferable to hide or remove certain sensitive or potentially identifying data elements (such as name, address, birthdate, age, disability, race or gender) for crimes such as domestic violence, sexual assault or stalking, rather than remove the crime incident altogether from the data set. A full data set, with sensitive or identifying data elements hidden or removed, can help provide a reliable estimate of the volume, frequency, and scope of the crime in a specific community.

Victim Names: It is unnecessary for open data sets to contain names of individual victims or witnesses. Particularly for sensitive crimes, such as domestic violence or sexual assault, open data sets should not include the names of victims and witnesses, even where the public record laws may not prohibit such disclosure. Victims' names -- even witness names, if they are family members or neighbors -- are identifying and could inadvertently reveal the identity of the victim and could result in backlash or unintended harm.

Suspect Names: Because of the intimate nature of domestic violence, sexual assault, and stalking crimes, agencies should also be cautious about publishing perpetrators' names, since knowing the perpetrator's identity could reveal the victim's identity. In addition, people initially arrested as suspects in domestic violence cases sometimes turn out to in fact be more properly classed as victims. In these cases, publishing their names may result in further victimization.

Location: Because domestic violence and sexual assault often occur in the victim's home, school, or place of work, the incident location could identify a victim. Agencies should not publish exact location, whether it is the full address or specific geographic coordinates (longitude, latitude).[i] Depending on the community, a block address in a densely populated area may be sufficient to mask exact location; however, in less populated areas in which a block has few houses or in locations that have small numbers of individuals with certain demographics, even a block address could be identifying. In these circumstances, location can be classified at a higher level of geography, such as neighborhood, police district, census tract, etc. without losing the incident-level details. For example, in a rural community where only 2-3 houses are on a block, the location data could be of the police district rather than the block address.

Another option is to provide location data in a table separate from other details, such as demographics and crime type, while limiting the ability for the data sets to be combined and re-identified.

This project was supported by Grant No. 2013-TA-AX-K006 awarded by the Office on Violence Against Women, U.S. Department of Justice (USDOJ OVW). The opinions, findings, conclusions, and recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the USDOJ OVW. If you have questions, please email: safetynet@nnedv.org

Combination of Identifiers: Even if no overtly personally identifying information is posted, a combination of demographics data[ii] could still inadvertently reveal a specific person as being a victim of domestic violence, sexual assault, or stalking. The combination of identifiers might include location, age, gender, race, ethnicity, or other demographics. For example, if the data set includes these elements: rape of a minor, victim's age is 12, victim's gender is female, and occurred on a specific block – this could be identifying if there is only one girl or very few girls of that age living on that block. Additionally, information from the data set could, in combination with other external data sets, create a "mosaic effect" where the combination of data can lead to re-identification of a victim.

A method to reduce the ability for someone to be identified through a combination of demographic data is to restrict demographic details for "outliers,"[iii] similar to how the U.S. Census Bureau publishes census data. This restriction is particularly necessary when the location has few individuals that fit a specific demographic. For example, if a jurisdiction serves a community which includes few African Americans, the data set should hide race demographics and publish only other data elements that would not be potentially identifying, such as date and time of the crime, type of crime, etc. Agencies should consider their community make up, how the data elements in their data sets could potentially reveal someone's identity, and take steps to remove certain data elements to minimize identifying a victim.

Narratives: Some data sets contain narratives describing the crime or interaction between law enforcement and offender. Narratives, which are also sometimes called "freeform" or "unstructured" fields, can contain details that could be potentially identifying, even if names are excluded. Agencies may choose to *rewrite* narratives before publishing to ensure that they are not identifying. It is advisable to *remove* narratives for sensitive crimes, such as domestic violence and sexual assault. If publishing the narratives as-is, agencies should institute a pre-publication review process for all narratives for sensitive crimes to ensure that they don't inadvertently reveal the victim's identity, keeping in mind that even de-identified narratives, when combined with the demographic data, could reveal a victim's identity. If a jurisdiction lacks the resources to conduct a review process, then the data should not be available through open data sources.

Delay Publication of Data: Another method to minimize the sharing of potentially identifying information is to delay publishing data sets. In general, police data do not have to be published immediately. Delayed publication will give the agency time to remove inaccurate data, review the data for potential re-identification, and ensure that what is being published adheres to agency policies. Additionally, a delay in release of the data may decrease risk in some cases.

Complementary Approaches to Sharing Incident Details While Minimizing Privacy Risks

The methods above primarily describe ways to redact individual data elements or classify data elements to a larger category to protect victim privacy. However, by removing these data elements, important information for advocates and decision-makers may be lost. Here are some complementary data sets

This project was supported by Grant No. 2013-TA-AX-K006 awarded by the Office on Violence Against Women, U.S. Department of Justice (USDOJ OVW). The opinions, findings, conclusions, and recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the USDOJ OVW. If you have questions, please email: safetynet@nnedv.org

that could be published alongside the redacted incident-level data to provide the community with insights they need for effective advocacy.

Aggregation of Variables: For certain types of data variables, such as age, race, or gender, aggregate data will be most protective of victim privacy and prevent re-identification. Aggregate data may not always provide the full context of a specific case, but could be very useful for identifying trends of certain crimes in a community and patterns in the law enforcement response to those crimes. When published alongside incident-level data where victim demographics or location has been redacted to protect privacy, aggregate data can provide crucial context and analysis.

Aggregation of domestic violence and sexual assault data could include: release of a separate dataset on domestic violence or sexual assault cases that provides information on the month and year of the crime, rather than the specific date/time; information on gender and broad age ranges aggregated to a level which protects anonymity, rather than more specific victim/offender information; race and ethnicity data at a location level that protects anonymity; or location data for the census tract or police district, rather than the exact address or block.

Identifying Data or Details for Research: In cases where an identifying data set is needed for research, the detailed data set could be made available for qualified researchers. Individuals with access to these data sets may require approval from an Institutional Review Board (IRB) and should establish agreements (e.g., Data Use Agreements or Memoranda of Understanding) affirming that they will not share data in a manner that could jeopardize confidentiality. A Privacy Certificate, for example, is a requirement for some federally-funded research, serving as an acknowledgement that the researchers understand their legal obligations to protect identifiable data.

Additional Reading

- NNEDV, “Issue Summary: Police Data Initiatives and Domestic Violence/Sexual Assault Victims”
- NNEDV, “How Law Enforcement Agencies Releasing Open Data Can Protect Victim Privacy & Safety”
- NNEDV, “Why Privacy & Confidentiality Matters to Victims of Domestic Violence and Sexual Assault”
- US DOJ, “Identifying and Preventing Gender Bias in Law Enforcement Response to Sexual Assault and Domestic Violence”
- President’s Task Force on 21st Century Policing. 2015. “Final Report of the President’s Task Force on 21st Century Policing”. Washington, DC: Office

[i] Geographic coordinates are easy to reverse-engineer, so removing addresses but leaving coordinates does little to protect privacy. Furthermore, adding noise to geo coordinates is not a good approach because it creates the potential for false re-identification (when the wrong person is associated with information in open data).

[ii] http://dataprivacylab.org/projects/identifiability/paper1.pdf

[iii] https://www.census.gov/srd/papers/pdf/rrs2004-03.pdf

This project was supported by Grant No. 2013-TA-AX-K006 awarded by the Office on Violence Against Women, U.S. Department of Justice (USDOJ OVW). The opinions, findings, conclusions, and recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the USDOJ OVW. If you have questions, please email: safetynet@nnedv.org