How Law Enforcement Agencies Releasing Open Data Can Protect Victim Privacy & Safety

During the past few years there has been increased public interest in accessing police data, and governments and law enforcement agencies have responded by making police data more broadly available in a variety of ways, including in some instances by publishing it online. Numerous police jurisdictions have published or are considering publishing police data online as part of open government data initiatives or projects[i] in order to improve transparency, provide the public with increased access to data, and provide a better understanding of their local law enforcement agency's practices, policies, and day-to-day operations. "Open data" is a term that is used to refer to incident-level data sets that are published online for public access in a format that permits them to be downloaded in their entirety and searched, sorted and analyzed.

Open police data can help inform the public about crime reports and how law enforcement responds to it, potentially revealing gaps and presenting opportunities to improve policing. The Department of Justice guidance on *Identifying and Preventing Gender Bias in Law Enforcement Response to Sexual Assault and Domestic Violence*[ii] and recent civil rights investigations of police departments' handling of domestic violence and sexual assault cases, for example, have demonstrated how police data, in combination with other data sources, can illuminate key issues and responses within the law enforcement and criminal justice systems and can be used to improve police training, protocols and monitoring to better support victims.[iii] Because ongoing efforts in many jurisdictions to improve the response to crimes of domestic violence, sexual assault and stalking may result in more reporting, it is important to note that an uptick in reports does not necessarily reflect an increase in incidence of these offenses. Rather, it might indicate a greater willingness among victims to go to law enforcement for help.

On the other hand, it is important to keep in mind that for victims of domestic violence, sexual assault, and stalking, their privacy is often fundamentally linked to their safety. Since some police data are published online and freely available to the public, it is critical to ensure that these data sets are presented in a way that does not facilitate their misuse. Granular, incident-specific data can make victims of crimes easily identifiable, which in turn can make them more vulnerable to further trauma, harassment, and discrimination in their personal or professional lives. As law enforcement agencies consider making more data open and available to the public, they have ethical and legal obligations to protect victim privacy.

In addition, even the perception of lack of privacy may prevent victims from reaching out for help. In fact, in a National Domestic Violence Hotline survey[iv] of callers regarding law enforcement responses, approximately 51% of victims who contacted the Hotline said they had not called the police, and, of those respondents, 60% said they did not call the police for privacy reasons. Many survivors experience discrimination or harassment for being victims of domestic or sexual violence or may face retaliation for reporting the crime. They may risk being evicted from their homes, dismissed from their jobs or face public skepticism and blame for the abuse they have suffered. Furthermore, victims may feel even more traumatized when their personally identifying information and details about an assault are published

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

online. For these reasons, law enforcement has a responsibility to safeguard the privacy of victims of domestic violence, sexual assault, and stalking to the greatest extent possible.

In determining whether and how to make larger amounts of data open to the public, law enforcement agencies should carefully consider which data to publish. This requires balancing the potential value of open data with potential negative consequences of sharing certain data, in order to find mechanisms that increase transparency of law enforcement responses while protecting victim privacy and confidentiality.

Below are several suggestions that could assist in developing those policies and practices.

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).[v] 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[vi] 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,"[vii] 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 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

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

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.

OTHER STEPS TO PROTECT VICTIM PRIVACY

Work with Victims

Law enforcement agencies should adopt a victim-centered approach to data privacy. The risks to a victim's privacy and safety may vary drastically from one case to another depending on the nature of the crime, the motivations of the suspect/offender and other factors. A victim often has valuable insight into the level of risk s/he faces. Ideally that input should be sought and a process should be in place to flag data in higher risk situations.

Work with Community Members

Law enforcement agencies should also consider working with their communities, specifically local domestic violence and sexual assault organizations as well as their state domestic violence and sexual assault coalitions, to determine how open data could be helpful in assessing these issues in their communities and how victim privacy can be protected. Other organizations working with specific populations, such as LGTBQ, immigrant, culturally specific communities, etc., could also be helpful in identifying how open crime and policing data could be shared and used in a responsible and beneficial way.

Review State Privacy and Open Records Laws

Some states have specific laws that protect victim privacy. Review these laws and reassess open records laws to determine whether publishing police data online may inadvertently violate these laws. Some states have restrictions on certain types of demographic data, such as age of offender or victim (for example, in most cases, juveniles are not named or their names are redacted), restrictions based on type of crime (for example, names of sexual assault victims are redacted), or restrictions on sharing information if disclosing the record would hinder an investigation or put someone at risk. Depending on the state, other exemptions may apply.

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

Even where the public records laws may allow disclosure, jurisdictions should consider whether, due to the sensitive nature of the information, individuals seeking access to such detailed information should be required to go through the traditional process of requesting and obtaining those records with proper redactions rather than automatically obtaining access to the records of significant volumes of cases through an online open data set. Jurisdictions may also want to consider updating the public records laws to reflect the significant changes due to the development and availability of online open data sets that were not even contemplated when public record laws were enacted.

Additional Reading

- President's Task Force on 21st Century Policing. 2015. "Final Report of the President's Task Force on 21st Century Policing". Washington, DC: Office
- US DOJ, Identifying and Preventing Gender Bias in Law Enforcement Response to Sexual Assault and Domestic Violence[viii]
- NNEDV, "Issue Summary: Police Data Initiatives and Domestic Violence/Sexual Assault Victims"
- NNEDV, "Why Privacy & Confidentiality Matters to Victims of Domestic Violence and Sexual Assault"

[i] https://www.whitehouse.gov/blog/2016/10/13/growing-number-communities-are-using-data-improve-policing-and-criminal-justice

[ii] https://www.justice.gov/opa/file/799366/download - Principle 8 encourages law enforcement agencies to "Maintain, review and act upon data regarding sexual assault and domestic violence."

[iii] http://sunlightfoundation.com/blog/2016/08/31/using-data-to-track-police-response-to-sexual-assault/

[iv] http://www.thehotline.org/wp-content/uploads/2015/09/NDVH-2015-Law-Enforcement-Survey-Report.pdf

[v] 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).

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

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

[viii] https://www.justice.gov/opa/file/799366/download

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