



# Open Data and Policing

## A Five-Part Guide to Best Practices

### **PART I: DEVELOPING OPEN DATASETS**






# Open Data and Policing

## A Five-Part Guide to Best Practices

### **PART I: DEVELOPING OPEN DATASETS**



This project was supported by Cooperative Agreement Number 2016CRW XK001, awarded by the U.S. Department of Justice, Office of Community Oriented Policing Services. The opinions contained herein are those of the author(s) and do not necessarily represent the official position or policies of the U.S. Department of Justice. References to specific agencies, companies, products, or services should not be considered an endorsement by the author(s) or the U.S. Department of Justice. Rather, the references are illustrations to supplement discussion of the issues.

The internet references cited in this publication were valid as of the date of this publication. Given that URLs and websites are in constant flux, neither the author(s) nor the COPS Office can vouch for their current validity.

The U.S. Department of Justice reserves a royalty-free, nonexclusive, and irrevocable license to reproduce, publish, or otherwise use and to authorize others to use this resource for Federal Government purposes. This resource may be freely distributed and used for noncommercial and educational purposes only.

Recommended citation:

Police Foundation. 2018. *Part I: Developing Open Datasets*. Open Data and Policing: A Five-Part Guide to Best Practices. Washington, DC: Police Foundation.

Published 2018





# Contents

I. Introduction . . . . .	1
II. What is an Open Dataset? . . . . .	2
III. What Type of Data Should Be Released? . . . . .	3
IV. Case Studies . . . . .	4
Portland (Oregon) Police Bureau . . . . .	4
Northampton (Massachusetts) Police Department . . . . .	7
V. Conclusion . . . . .	10
About the Police Foundation . . . . .	11





# I. Introduction

Law enforcement agencies nationwide, large and small, have already released open datasets representing citizen calls for service, arrests and citations, police response to resistance, assaults on officers, citizen complaints, bicycle and pedestrian collisions, traffic stops, drug overdoses, hate crimes, agency training and workforce data, and more. By providing open data, they allow members of the public, community groups, and law enforcement agencies to independently and collaboratively analyze the data to identify potential problems, improve understanding of the challenges faced by law enforcement and their responses, craft solutions, and improve their communities. To date, more than 140 law enforcement agencies have joined the [Police Data Initiative](#) (PDI), a community of practice consisting of members that have committed to releasing open data.

Drawing from promising practices used by these law enforcement agencies, this five-part best practice series aims to guide executives and members of local law enforcement agencies as they release open data. This guide, *Part I: Developing Open Datasets*, outlines key points for agencies to consider in developing their datasets for public release.

## II. What is an Open Dataset?

Various definitions of open data exist. Generally, open data is

- a set of “granular” information (i.e. incidents rather than aggregations or summaries);
- in a downloadable, digital format;
- available online at no cost;
- readily machine-readable and analyzable (i.e., no PDF, image, or other locked files);
- nonproprietary and without restriction;
- released by the original information source in a timely manner.<sup>1</sup>

For examples of open datasets, see the [PDI website](#). For more on the value of open data, see [Law Enforcement Executive’s Guide to Open Data: Supporting the Community in the Co-Production of Public Safety](#).

---

1. For more information, see Joshua Tauberer, “14 Principles of Open Government Data,” in *Open Government Data: The Book* (2nd ed.) (self-pub., 2014), <https://opengovdata.io/2014/principles/>.



### III. What Type of Data Should Be Released?

Once agencies decide to release open data, the first step is to determine the type of data to release. Selecting a dataset and fields for that dataset depends on a variety of factors for an agency and its community. Some main considerations include the following:

- **Data Availability:** Agencies are already required to report certain information to state and federal agencies and typically maintain records of their activities within internal systems. These systems may offer an easy means of translating information into open data.
- **Capacity to Compile and Release Data over the Long Term:** Agencies should develop a plan for dataset releases and updates that account for their records system and staffing and their desired frequency of dataset updates. For example, the **Louisville Metro (Kentucky) Police Department** found certain datasets, including [open data on hate crimes](#), relatively easy to prepare and make open since they already participate in the National Incident-Based Reporting System (NIBRS), a crime data collection program run by the Federal Bureau of Investigation.
- **Police and Community Interests and Priorities:** Agencies should consider the value an open dataset would pose for the agency, the community, and learning opportunities more generally.
  - Partnering with their local library system, the **Chapel Hill (North Carolina) Police Department** released open data on incidents, traffic stops, employee demographics and more, in part to proactively provide members of the public information that they would regularly request.
  - The **Henderson (Nevada) Police Department** released open data on their use of force incidents with their internal year-end analysis on a webpage discussing use of force, providing the community with both the data and its context.
  - The **Ferndale (Michigan) Police Department** released open data on local drug incidents and responding officers' use of naloxone, a drug that can reverse an opioid overdose.
- **Sensitive Information:** Datasets should make adjustments for potentially [sensitive information](#), such as information related to open or active cases that might compromise police investigations or other personally identifiable information (PII) that might compromise victim privacy and safety. For example, the **Ft. Lauderdale (Florida) Police Department** includes a random offset for privacy purposes, and removes locational data for certain crimes in accordance with state law.
- **Other Considerations:** Agencies should consider the presentation of their data for public consumption. One way that agencies can ensure they are developing open datasets with understandable data fields is to identify and engage the community to test a beta version of the open dataset prior to release. Based on feedback, agencies should be prepared to make necessary adjustments to their dataset design.

## IV. Case Studies

### Portland (Oregon) Police Bureau

*“Open data can improve transparency and create an informed community, but it is important to take the time to make sure the data you share is accurate and ethical.”*

– Captain Mike Krantz, Portland Police Bureau

The Portland Police Bureau (PPB) is the law enforcement agency for the city of Portland, Oregon. Over 900 sworn PPB officers serve the jurisdiction of approximately 640,000 people.<sup>2</sup> The Analysis Unit within the PPB’s Strategic Services Division leads the bureau’s open data effort. Crime Analysts within this unit are responsible for developing open datasets, interactive dashboards, and supporting materials. The PPB joined the PDI in September 2016.

In early 2015, the PPB switched to a new records management system and took the initiative to become compliant with the FBI’s NIBRS reporting. With a desire to respond to community stakeholders interested in crime information, PPB began the process of developing open datasets and prioritized monthly neighborhood offense information, defined as NIBRS Group A offenses reported to the PPB, as their first open dataset.

When they began developing the open dataset, the PPB considered a variety of stakeholder needs as they began to choose dataset fields and select the best means of presenting their data. Since Portland is a very neighborhood-centric city, the PPB knew it would be important to present the data at a level of analysis familiar to the community. They chose to develop an interactive dashboard through Tableau, which would provide visualizations of the data at the neighborhood level for those merely seeking summary information and which would provide an option to download the entire incident-level dataset for those seeking to analyze the data on their own.

---

2. “2016 Crime in the United States,” Federal Bureau of Investigation, accessed December 22, 2017, <https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/tables/table-26/table-26-state-cuts/table-26-oregon.xls>.

**Figure 1. Portland Police Bureau steps to developing open datasets for release**



The NIBRS data points in the PPB’s planned dataset on reported crimes were validated through the RMS, verified through NIBRS checks, and reviewed for accuracy during record processing. The team worked with personnel from different areas of the agency—command staff, records management staff, the public information team, neighborhood officers—to build bureau-wide buy in and develop internal stakeholder engagement. The data team wanted to ensure that the process of opening data was transparent and inclusive and incorporated a collaborative approach that would benefit all involved without putting any case or victim at risk. For example, PPB decided to include real case numbers (on non-sensitive cases) as a data field because the data team worked with the Records Division to determine that providing that information could help to reduce the number of public records requests for case numbers that Records receives each year.

The data team consulted with command staff, public records personnel, and the City Attorney’s office to develop solutions that would balance the value of sharing information with the need to protect the privacy of individuals. As such, the PPB internally developed a process whereby all address locations in the data would be rounded to the nearest 100-block rather than identify the specific address. Additionally, records in the RMS that have been marked sensitive by investigators are automatically stripped of address data and are only provided at the neighborhood level. This approach is applied to all cases involving sexual assault, domestic violence, or a juvenile victim.

Once the dataset was ready for review, the data team checked with PPB command staff for final permissions to publish the data. Before sharing broadly with the community, the team tested a beta version of the dashboard with internal users and provided a demonstration of the dashboard to external crime prevention partners who work

closely with community members. The feedback from these user groups informed the creation of supporting documents, including a walkthrough and FAQ section, to help users better understand and use the dashboard.

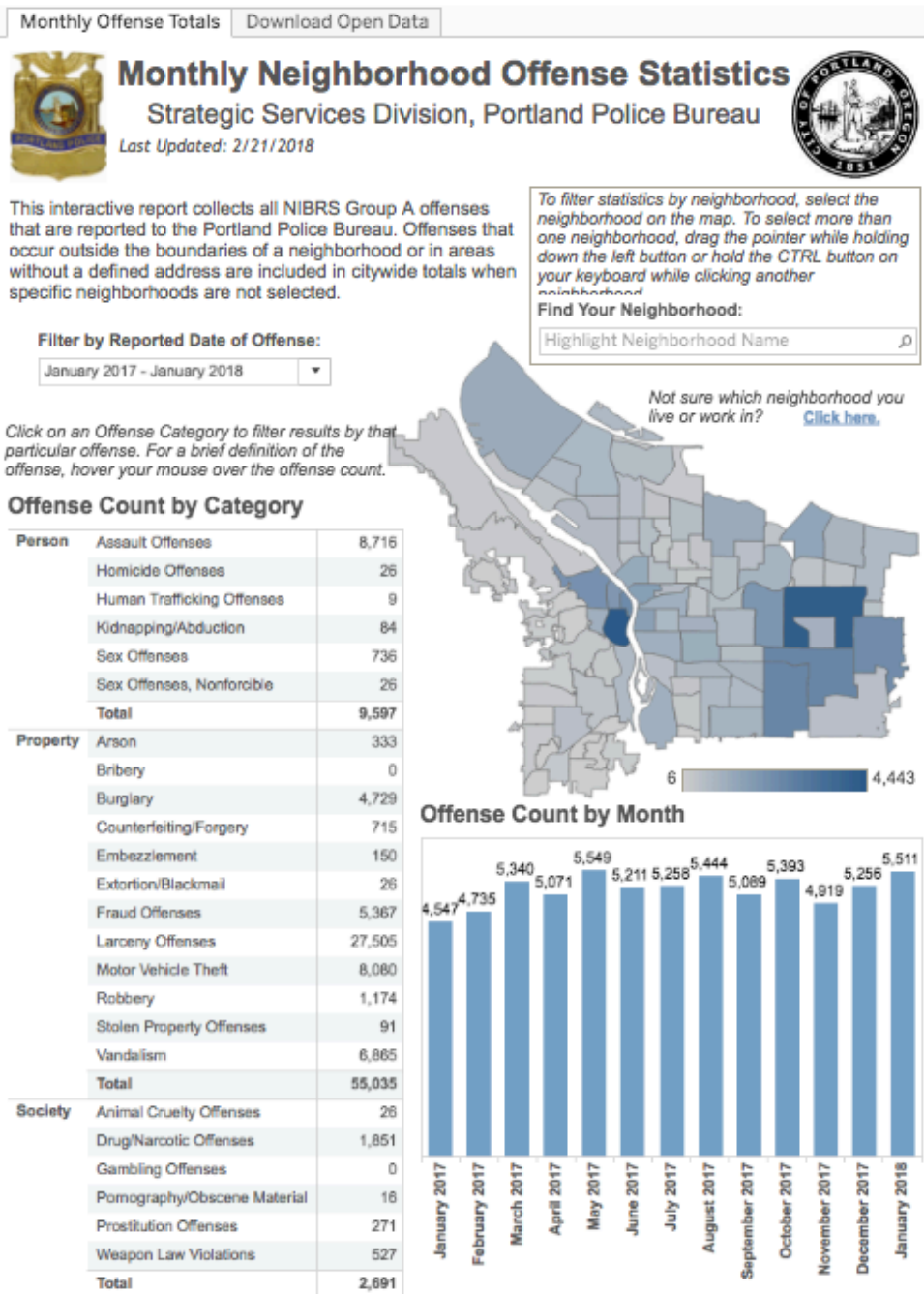
## Portland Police Bureau Open Data Dashboard

## Crime Statistics

**Due to record processing delays, February 2018 data will be updated by March 26th, 2018.**

**This report is updated the first business day after the 20th of the month.**

To download the most recent data set powering this report, click on the gray "Download Data" tab and follow the instructions on the page.



When the dashboard went live in April 2017, the PPB used social media to promote the launch. The team monitored social media feedback following the release and incorporated it into updates, such as releasing additional fields in the open dataset. The public feedback and the PPB's willingness to revise the dashboard and dataset were invaluable to the tool's successful use and continued maintenance.

## Northampton (Massachusetts) Police Department

*"One positive outcome [of releasing open data] is building trust and legitimacy. But it's not the only benefit."*

— Chief Jody D. Kasper, Northampton Police Department

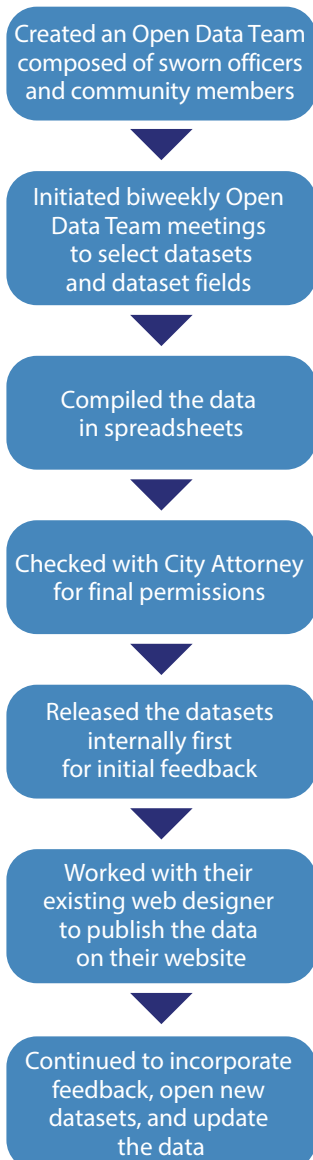


Northampton  
(Massachusetts) Police  
Department badge

The Northampton Police Department (NPD) is the law enforcement agency serving the City of Northampton, Massachusetts. The department employs 65 officers serving approximately 28,500 residents in their jurisdiction.<sup>3</sup> The NPD joined the PDI as a participating agency in May 2016 in an effort to increase transparency and build community trust. Following the commitment, NPD Chief Jody D. Kasper's first step was to pull together an Open Data Team (ODT). The Open Data Team is composed of sworn officers and volunteer residents of the City of Northampton. The team met biweekly during the summer of 2016 to design the NPD's open datasets.

3. "NPD Employee Demographics, Race/Ethnicity," Northampton Massachusetts Police Department, accessed January 12, 2018, <https://www.northamptonpd.com/npd-open-data-portal/npd-employee-demographics/233-npd-employee-demographics-race-ethnicity.html>; "QuickFacts: Northampton city, Massachusetts," U.S. Census Bureau, accessed January 12, 2018, <https://www.census.gov/quickfacts/fact/table/northamptoncitymassachusetts/PST045216>.

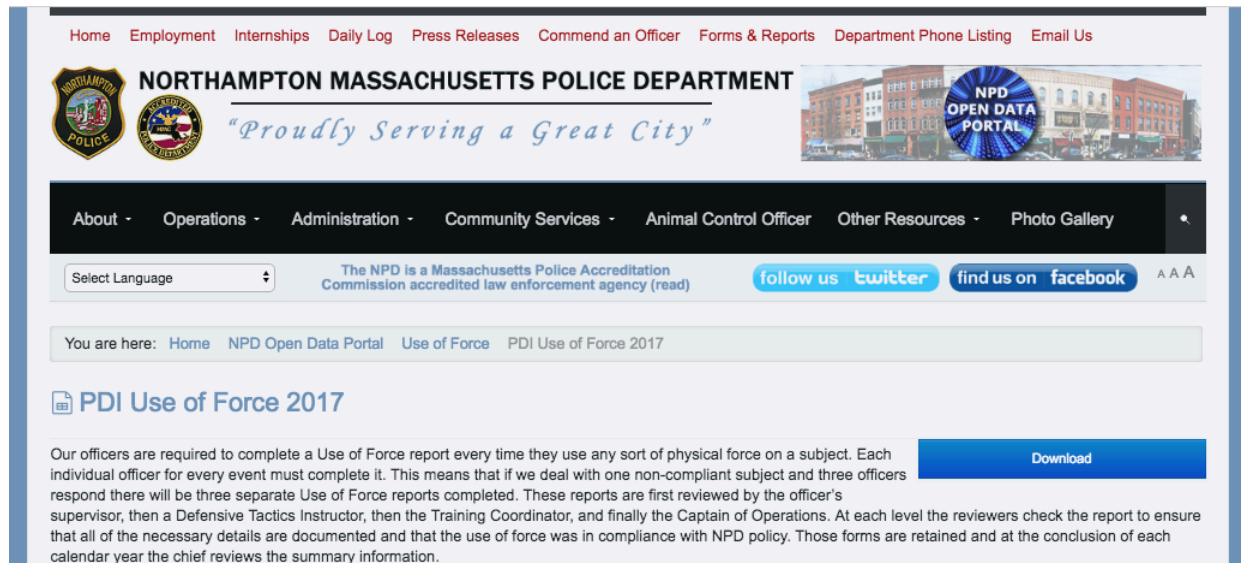
**Figure 2. Northampton Police Department steps to developing open datasets for release**



The ODT began to create data tables—including use of force, employee demographics, and officer-involved shootings—that the team initially identified as being of interest to their community, and that would be relatively easy to prepare. The team prioritized datasets they believed would provide useful information for their community to know. To select data fields, the ODT used examples of open data tables from other agencies that have similar categories of data featured on the PDI website.

The ODT made recommendations during group discussions. The sworn members of the ODT mined the data, created the tables, and obtained feedback from the larger ODT. As a smaller agency, the NPD lacked the funding to hire outside help for data mining, so ODT members created and populated tables. Interns and support staff also helped with some of the dataset compilation. Given the manual input required and to establish consistency, the team decided to post all future dataset updates on an annual basis at the end of each calendar year. Ultimately, a mix of 10 to 15 open and summary datasets were selected for release and sent to the city attorney for review. These datasets were also preemptively released within the department for initial feedback. The NPD published their approved datasets on the NPD website as downloadable Excel files. Each dataset was given its own page with an explanation of the dataset.

Figure 3. Northampton Open Data Portal



As the first police department in Massachusetts to release open data, Chief Kasper felt the NPD received an immediate positive public response. Since the initial release, the NPD has continued to open more datasets and update their data. The quick access that open data provides has helped the department to more efficiently tackle tasks like grant writing, press interviews, and public forums, where having on-hand statistical data is useful. Likewise, the NPD is able to respond to public record inquiries much more efficiently as they direct requesters to the website where the open data is published. This has greatly eased the department's administrative workload. Instead of spending significant time and effort to compile requested information on a case-by-case basis, such as information on use of force, after releasing open data the NPD could "send them a website link to the open data portal, [and] say 'it's all right here, go for it.'"<sup>4</sup> Furthermore, the preemptive release of data has allowed the department to guide the narrative behind police and crime trends with firsthand insights.

4. Jody D. Kasper, Northampton Police Department, interview with Police Foundation staff, January 12, 2018.



## V. Conclusion

As law enforcement agencies take steps to use open data, those involved in the open data process should take care to develop their datasets with planning and engagement. Developing a formal departmental strategy for open data releases that accounts for a variety of valuable design factors will ease the open data release process. For example, the City of Portland has adopted an ordinance to establish a citywide open data policy. Key lessons learned from Portland, Northampton, and other PDI agencies that have successfully developed open datasets are as follows:

- Develop a plan for open data releases.
- Consider data availability.
- Consider agency capacity.
- Consider community interests and priorities.
- Balance privacy concerns.
- Engage members of the department and community in the development process.
- Continue to update data and incorporate feedback as appropriate.



## About the Police Foundation

The **Police Foundation** is a national, nonpartisan, nonprofit organization dedicated to advancing innovation and science in policing. As the country's oldest police research organization, the Police Foundation has learned that police practices should be based on scientific evidence about what works best, the paradigm of evidence-based policing.

Established in 1970, the foundation has conducted seminal research in police behavior, policy, and procedure and works to transfer to local agencies the best new information about practices for dealing effectively with a range of important police operational and administrative concerns. Motivating all of the foundation's efforts is the goal of efficient, humane policing that operates within the framework of democratic principles and the highest ideals of the nation.

To learn more, visit the [Police Foundation](#) online.





Law enforcement agencies nationwide have released open datasets representing calls for service, arrests, and more. Members of the public, community groups, and law enforcement agencies can analyze this data to identify problems and craft solutions. To date, more than 140 law enforcement agencies have joined the Police Data Initiative (PDI). Drawing from their promising practices, this five-part series aims to guide executives and members of local law enforcement agencies as they release open data.

This guide, *Part I: Developing Open Datasets*, outlines key points for agencies to consider in developing their datasets for public release. It covers the types of data that should be released, and contains case studies of police open data programs in Portland, Oregon, and Northampton, Massachusetts.



1201 Connecticut Ave NW  
Suite 200  
Washington, DC 20036

[www.policefoundation.org](http://www.policefoundation.org)

031801865  
Published 2018