INFX 551 Data Curation
January 23, 2017
Jacob Kovacs

Profile & assessment of Seattle Police Department's 911 Incident Response and Police Report Incidents data

Description of data products

In cooperation with the City of Seattle and Seattle's Department of Information Technology (SPD, 2017), the Seattle Police Department (SPD) has opened up some of its crime data through Seattle's open data portal, data.seattle.gov. As a result of insufficient contextual information, to understand one dataset I ended up profiling two: 911 Incident Response data (SPD, 2017b) and Police Report Incidents data (SPD, 2017d). Both datasets are findable through the open data portal and SPD's website (SPD, n.d. c), but neither access point makes clear exactly *what* this data is—how it is generated, how it is used by the data authors and owners, or how the two datasets differ.

Within the open data portal, there are "primer" pages with metadata for each dataset (SPD, 2017a; SPD, 2017c). Combining the primers' narrative descriptions (Box 1, Box 2) and attached public release rules (Table 1, Table 2) with a third source (CrimeStar, n.d.), my best guess is that the 911 response data is input to a computer-aided dispatch (CAD) information system by 911 operators, then updated with information from police officers once they've investigated the incident. The police reports are perhaps triggered by a civilian filing a complaint in-person or by an officer initiating an encounter, rather than by phone calls. This data is entered in an internal records management system (RMS). After processing and redaction of sensitive information, both CAD (911 incidents) and RMS (police reports) data are published online.

Box 1. Dataset description, 911 incident response data (quoted from SPD, 2017a).

This dataset is all the Police responses to 9-1-1 calls within the city. Police response data shows all officers dispatched. To protect the security of a scene, the safety of officers and the public, and sensitive ongoing investigation, these events are added to the data.seattle.gov only after the incident is considered safe to close out. Data is refreshed on a 4 hour interval.

Table 1. Public rules for CAD data release (quoted from SPD, n.d. a).

Officers close calls in the 911 system when the preliminary investigation is complete. Once the 911 call is closed the data will be released.
Narratives, remarks, text, entities, and descriptions may contain personal, juvenile, and national security information and are not released.
Each crime type is categorized by the specific MIR code, as well as by a generic and summary type for the ease of reporting.
911 Call data will be withheld if it contains personal, juvenile, or national security information.
Only CAD calls with valid final MIR codes will be released.
911 data is always updated with the latest information.
Only the primary 911 call for each event is released for the ease of reporting.

Box 2. Dataset description, police report incidents data (quoted from SPD, 2017b).

These incidents are based on initial police reports taken by officers when responding to incidents around the city. The information enters our Records Management System (RMS) and is then transmitted out to data.seattle.gov. This information is published within 6 to 12 hours after the report is filed into the system.

Table 2. Public rules for RMS data release (quoted from SPD, n.d. b).

Initial Case Information is released after the report is transcribed providing no other exclusions apply.
Case offense codes are categorized by type and summary type for the ease of reporting.
Cases which include offense codes for personal, juvenile, and national security crimes are not released.
Care records that were/are under court order and sealed, retracted or expunged are not released.
Cases occurring at addresses which contain personal, juvenile, and national security information are not released.
Initial Case Data is available after the police report has been transcribed. Generally this takes 2 days.

The CAD dataset so large it can't be opened in Excel—the analytic tool people are most likely to have available. When accessed, the CAD dataset contained about 1,363,000 rows (SPD, 2017b) versus Excel's maximum of 1,048,576 rows (Microsoft, n.d.). The RMS dataset was slightly smaller with about 863,000 rows. For both datasets, in addition to in-browser tabular and map views of the data (provided by Socrata software), there are options to access the data through an API (application programming interface); to create a refreshable Excel file using OData; or to download the data as a CSV, JSON, RSS, RDF, TSV, or XML file.

Each dataset has 19 fields (columns), but there is no metadata explaining each field's contents. In fact, three out of eight comments in the discussion section accompanying the CAD dataset (SPD, 2017b) and twelve out of seventeen comments in the discussion section accompanying the RMS dataset (SPD, 2017d) are complaints about difficult interpreting the data due to insufficient metadata. People posed questions about apparent record duplication, apparent and unexplained exclusion of rape as a crime category, how dates and times were encoded, what codes represented, etc.; most questions were two or three years old but had received no answer.

Again making my best guess, the CAD dataset seems to contain about two unique record identifiers, six category codes for reporting purposes, two narrative description fields, two date/time fields, and seven location-related fields (Table 3); perhaps fields named "clearance" are based on a final officer-authored report, while fields named "initial" are input by 911 dispatchers. Similarly, the RMS dataset seems to contain one unique identifier, five category codes, one narrative description field, five date/time fields, and seven location-related fields (Table 4).

Table 3.

CAD CDW ID	Number
CAD Event Number	Number
General Offense Number	Number
Event Clearance Code	Number (code)
Event Clearance Description	Text
Event Clearance SubGroup	Text (code)
Event Clearance Group	Text (code)
Event Clearance Date	Date and time (mm/dd/yyyy hh:mm:ss AM/PM)
Hundred block location	Text (cross streets)
District/Sector	Text (code)
Zone/Beat	Text (code)
Census Tract 2000	Number (code)
Longitude	Number
Latitude	Number
Incident Location	Number (latitude, longitude)
Initial Type Description	Text
Initial Type Subgroup	Text (code)
Initial Type Group	Text (code)
At Scene Time	Date and time (mm/dd/yyyy hh:mm:ss AM/PM)

Table 4.

RMS CDW ID	Number
General Offense Number	Number
Offense Code	Number (code)
Offense Code Extension	Number (code)
Offense Type	Text (code)
Summary Offense Code	Number (code)
Summarized Offense Description	Text
Date Reported	Date and time (mm/dd/yyyy hh:mm:ss AM/PM)
Occurred Date or Date Range Start	Date and time (mm/dd/yyyy hh:mm:ss AM/PM)
Occurred Date Range End	Date and time (mm/dd/yyyy hh:mm:ss AM/PM)
Hundred Block Location	Text (cross streets)
District/Sector	Text (code)
Zone/Beat	Text (code)
Census Tract 2000	Number (code)
Longitude	Number
Latitude	Number
Location	Number (latitude, longitude)
Month	Number (code)
Year	Number (yyyy)

Evaluation of potential for reuse

Over the lifetime of the site (2/3/2013-1/23/2017), the 911 incidents dataset is second most popular at 110,395 views, and the police reports dataset is fifth most popular at 65,980 views (City of Seattle, 2017). If viewing is a form of reuse, then these datasets have a comparatively strong track record.

Most of the credit for these impressive numbers should probably go to the user communities that seek out the data, rather than to the curators of the open data portal. Many parties have a large stake in crime data (Lord, n.d.): homeowners, business owners, and real estate agents when weighing the merits of different properties; academic or professional researchers using crime data as a predictor or outcome in any number of questions; activists concerned about neighborhood safety or police activity; journalists seeking context for a story; individuals trying to find information about a specific crime they were involved in; etc.

With so many motivated groups, it's unsurprising that these datasets are popular despite major barriers to reuse. As noted above, metadata is inadequate. This not only hinders reuse, it promotes misuse—for instance, people will form opinions and make arguments on the basis of data they don't realize is partial (subject to unspecified filtering). Another barrier to reuse is that—although it's not made clear by the metadata—the datasets seem to follow local data standards. While there are a suite of national standards for data exchange between law enforcement agencies (UCR, N-Dex, NEIM, GJXDM) as well as an apparent open source effort (SPOTS) to standardize crime data for open data initiatives (FBI, n.d.; Suszan, 2014), it doesn't appear that a widely-accepted standard yet exists for this type of data.

At very least, this data is bound to be persistent. There are strong internal reasons to capture and preserve the data: it's used to inform police operations and to satisfy federal reporting requirements. There are also large external constituencies for crime statistics that—given current trends like Black Likes Matter (BLM, n.d.), the local Department of Justice ruling (DOJ, 2015), and Seattle's homelessness epidemic (Kroman, 2016)—are likely to grow larger and more desirous of data on crimes and policing.

Description and comparison of data portal

As mentioned, I accessed these datasets through the City of Seattle's open data portal. The portal is maintained by the Seattle Department of Information Technology in fulfillment of an executive order by Mayor Edward Murray (2016); the IT department offers the technical expertise required to support other city departments in opening up their data.

The portal is fairly navigable. From the homepage, the 911 incidents and police reports datasets can be quickly located by browsing the linked "Public Safety" category in the data catalog. Search is also available directly from the homepage. I did find the mixture of different content on the homepage confusing. They seem to be trying to showcase a lot of different kinds of resources (blog posts, how-tos, datasets, news items, etc.), and it distracts from actually accessing the data. I would recommend simplifying the homepage so that the main purpose of the site—to host open data—is immediately clear to visitors.

The portal is driven by Socrata software. The resulting capability to skim, map, and filter data in-browser is a major strength of the site. It makes basic data analysis much more accessible to site users. Still, lack of metadata and the need to learn the software interfere with users' ability to make knowledgeable use of the data. Currently, the "help" button on the nav bar leads to Socrata's generic support page, where explanations are more targeted for Socrata site administrators than confused citizens. I recommend that user-friendly Socrata tutorials be featured prominently on the site, and that metadata be adequate.

The portal has no data deposit process. Instead, users may suggest datasets. The opportunity to do so is promoted on the homepage with a button and at the bottom of several other pages with a link: “Didn't find what you're looking for? Suggest a dataset.” These links lead to a page where you can make a request or view others’ requests by status: all, open, approved, or rejected. While the interface is extremely user friendly, the actual requests have not been maintained. There have been about 90 requests over the lifetime of the site (seven years), with about 15 approved, 10 rejected, and the others simply ignored. I recommend that this backlog of requests be processed.

Finally, in contrasting this civic data portal with the Data Repository for University of Minnesota (DRUM, an institutional research-oriented data repository), a few interesting differences emerge. At a basic level, the same search and browse functionality exist, but there are many, many more subject categories in DRUM versus the city data portal (40,817 versus 10). This reflects the much wider range of topics that university researchers are concerned with, but it also makes the site much less navigable.

An additional point of difference is that DRUM, unlike the city data portal, relies on data submissions from users and therefore has a thorough data deposit process. DRUM foregrounds this data deposit process on its home page, along with many offers for assistance from library data experts—things like data management plan assistance, metadata consultations, and curatorial review of submissions (UML, n.d.).

Because DRUM data is intended for reuse by an expert audience of scholars, unlike the city data portal there are no services for analyzing the data through DRUM. Instead, users are expected to download datasets for analysis. They are able to reuse this data intelligently because thorough metadata is collected during the data deposit process.

References

Black Lives Matter (BLM). (n.d.). About the Black Lives Matter Network. Retrieved from http://blacklivesmatter.com/about/

City of Seattle. (2017, January 23). Site analytics. Retrieved from https://data.seattle.gov/analytics

CrimeStar. (n.d.). Police software - CAD vs. CFS. Retrieved from http://www.crimestar.com/cadvscfs.html

Department of Justice (DOJ). (2015, August 6). Investigation of the Seattle Police Department. Retrieved from https://www.justice.gov/crt/investigation-documents

Federal Bureau of Investigation (FBI). (n.d.). Uniform Crime Reporting. Retrieved from https://ucr.fbi.gov/

Kroman, D. (2016, June 29). Seattle's homeless emergency: No end in sight. *Crosscut.* Retrieved from http://crosscut.com/2016/06/seattles-homeless-emergency-no-end-in-sight/

Lord, J. (n.d.) Open data guide: Crime data. Retrieved from http://jlord.us/open-data-guide/crime/

Microsoft. (n.d.). Excel specifications and limits. Retrieved from https://support.office.com/en-us/article/Excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3

Murray, E. B. (2016, February 27). Executive Order 2016-01 [PDF file]. Retrieved from http://murray.seattle.gov/wp-content/uploads/2016/02/2.26-EO.pdf

Seattle Police Department (SPD). (2017a, January 21). 911 incident response. Retrieved from https://data.seattle.gov/Public-Safety/Seattle-Police-Department-911-Incident-Response/3k2p-39jp

Seattle Police Department (SPD). (2017b, January 21). 911 incident response [dataset]. Retrieved from https://data.seattle.gov/Public-Safety/Seattle-Police-Department-911-Incident-Response/3k2p-39jp/data

Seattle Police Department (SPD). (2017c, January 21). Police report incident. Retrieved from https://data.seattle.gov/Public-Safety/Seattle-Police-Department-Police-Report-Incident/7ais-f98f

Seattle Police Department (SPD). (2017d, January 21). Police report incident [dataset]. Retrieved from https://data.seattle.gov/Public-Safety/Seattle-Police-Department-Police-Report-Incident/7ais-f98f/data

Seattle Police Department (SPD). (n.d. a). Public rules for CAD data release [Microsoft Word file]. Retrieved from https://data.seattle.gov/api/assets/98E5052D-5E37-4BB4-B903-50BD73C97CCC?download=true

Seattle Police Department (SPD). (n.d. b). Public rules for RMS data release [Microsoft Word file]. Retrieved from https://data.seattle.gov/api/assets/BA4FF139-F37F-4CCA-871D-C827E8211DB0?download=true

Seattle Police Department (SPD). (n.d. c). SPD data sets. Retrieved from http://www.seattle. gov/seattle-police-department/crime-data/spd-data-sets

Suszan, B. (2014, March 12). The SpotCrime Open Crime Standard (SOCS). Retrieved from http://blog.spotcrime.com/2014/03/the-spotcrime-open-crime-data-standard.html

University of Minnesota Libraries (UML). (n.d.). Data Repository for U of M (DRUM). Retrieved from http://conservancy.umn.edu/handle/11299/166578