The Case for a National Repository of Policing Data in the United States

Cheryl Danton[1], Christopher Graziul[1], Elliott Ramos[2]

[1] University of Chicago; [2] CBS News

Open Repositories 2025
University of Chicago
Chicago, IL USA

Introduction

Policing data: Data collected or generated as part of policing activities

- Data repositories…
    - **Exist** for criminal justice data and court records
    - **Do not exist** for policing data!

- **Why is this a problem?**
    - Police transparency is difficult to achieve in practice (see Invisible Institute)
    - "Police Are Our Government" (Soss and Weaver 2017)
    - Members of the public have little insight into how their communities are policed!

What makes policing data so special?

- Focus on organizational process, not outcome
  - **Officer characteristics** (e.g., age, race/ethnicity, etc.)
  - **Officer behavior** (e.g., attendance, encounters with the public, etc.)
  - **Event descriptions** (e.g., when/what reported, involving whom, etc.)

- Sensitive information routinely included
  - **Accessible to the public**, often via legal statute
  - **Unique privacy risks**, rarely acknowledged

- Massive mess
  - **No standardization**, and Uniform Crime Reporting has major flaws (Comer et al. 2023)
  - **Collected *en masse***, and data quality/integrity (e.g., user error) can be questionable

What makes policing data so special?

Technical Details

- **Volume/type of records**
  - Millions of interactions with the public, daily
  - Hundreds of variables collected/tracked
  - Similar volume of unstructured data (e.g., radio transmissions, body worn cameras, etc.)

- **Inconsistent taxonomies**
  - Definitions are location-specific (e.g., "event")
  - No standardized data collection practices
  - Years/decades of incommensurate data

- **Lack of legal standards**
  - Public access through open records requests
  - Can contain sensitive personal information
  - Transparency vs. privacy tradeoff

Ethical Considerations

- **Data about people who...**
  - Are bystanders
  - Are victims
  - Are *not* subject to an enforcement action, or...
  - Are *not* convicted of a crime

- **Who owns the data? The public!**

- **What can be done with the data? A lot!**

- **Access controls**
  - Require new community-driven standards
  - Must still provide meaningful transparency
  - Who decides what is appropriate?

What would a National Repository of Policing Data enable?

Organizational analysis

- **Tabular data**
  - Who responds to what kind(s) of calls?
  - What are officers typically doing?
  - How are latent processes unfolding?
    - Overtime practices
    - Co-worker cliques and near-peer mentoring

- **Spatial data**
  - Where are officers typically concentrated?
  - What "spillover" effects shape deployment and/or response times?
  - How do events in one part of the city affect policing in other parts of the city?

- **Unstructured data**
  - What do officers learn before arriving on scene?
  - How do officers interact with members of the public?

Comparative analysis

- **Tabular data**
  - How do variable definitions differ across systems?
  - Why do variable definitions differ across systems?
  - What kind of comparisons are possible vs feasible?

- **Spatial data**
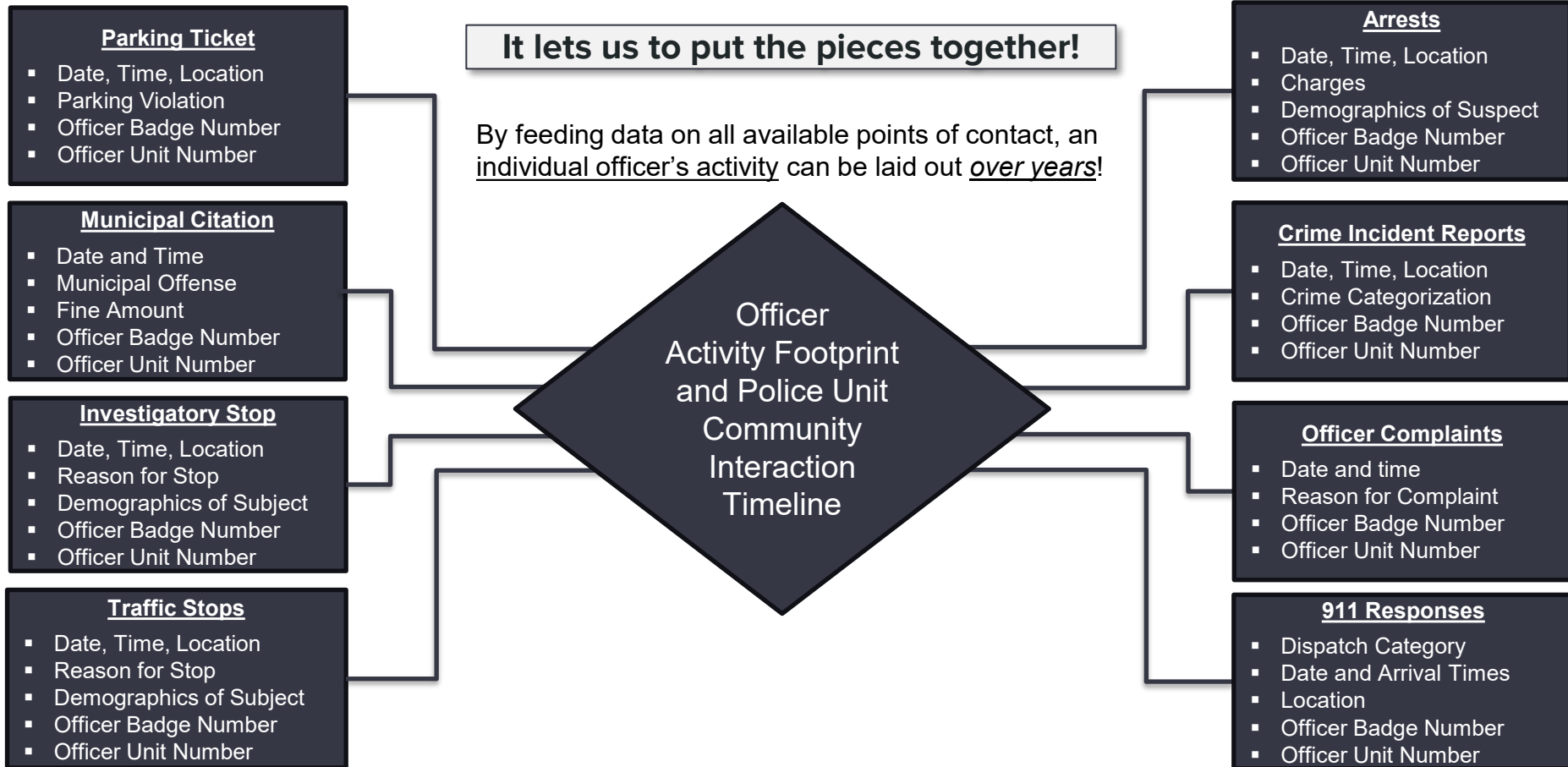 - Do police systems respond similarly to similar neighborhoods?
    - Same demographics, different city
    - Same city, different demographics
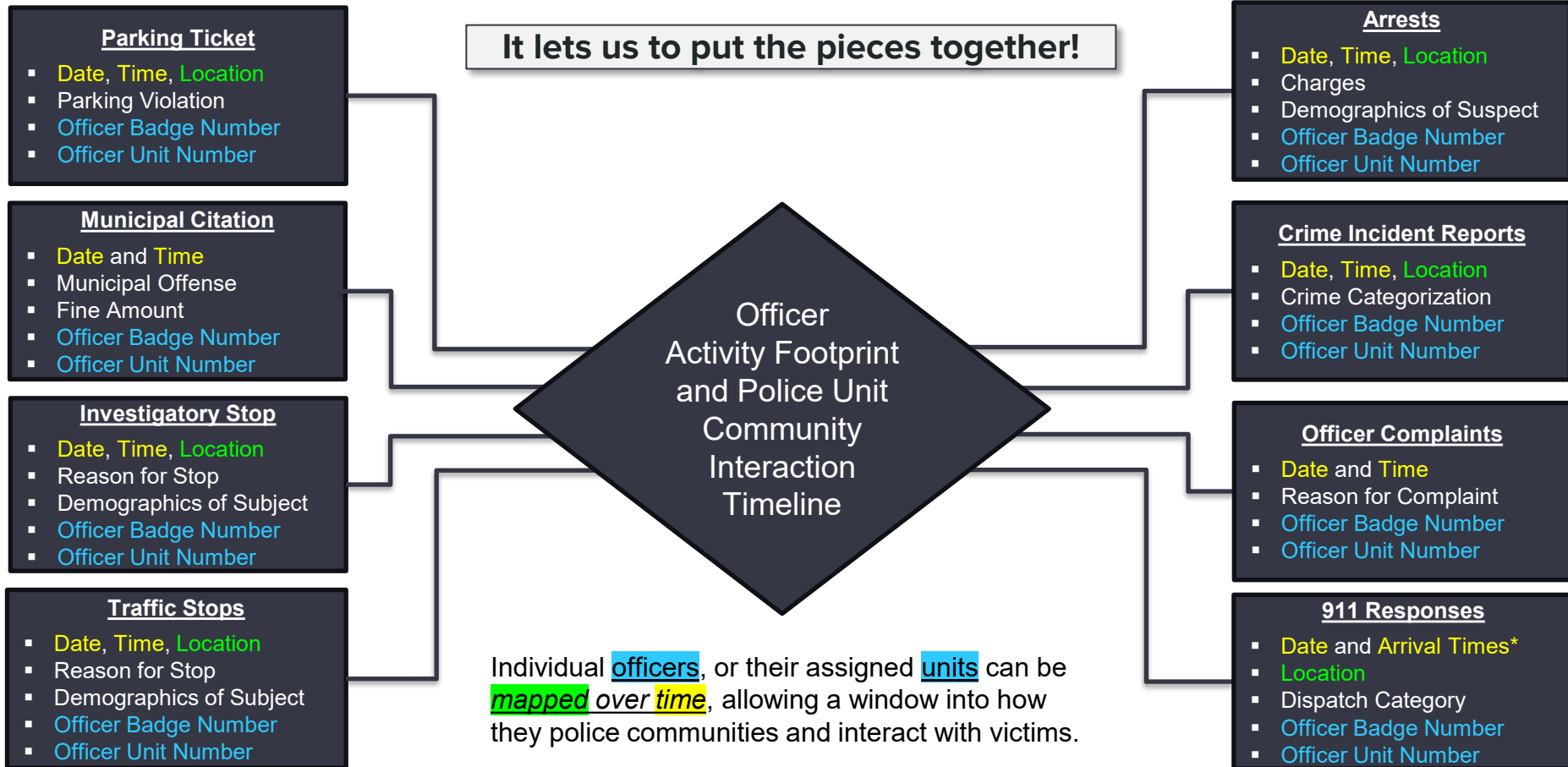 - Is "out of place" policing prevalent? If so, where?

- **Unstructured data**
  - How do communication practices shape the way different systems reply to similar situations?
  - What kinds of information (audio, video) are used or collected? Why?

What would a National Repository of Policing Data enable?

Parking Ticket
- Date, Time, Location
- Parking Violation
- Officer Badge Number
- Officer Unit Number

Municipal Citation
- Date and Time
- Municipal Offense
- Fine Amount
- Officer Badge Number
- Officer Unit Number

Investigatory Stop
- Date, Time, Location
- Reason for Stop
- Demographics of Subject
- Officer Badge Number
- Officer Unit Number

Traffic Stops
- Date, Time, Location
- Reason for Stop
- Demographics of Subject
- Officer Badge Number
- Officer Unit Number

It lets us to put the pieces together!

By feeding data on all available points of contact, an individual officer's activity can be laid out *over years*!

Officer Activity Footprint and Police Unit Community Interaction Timeline

Arrests
- Date, Time, Location
- Charges
- Demographics of Suspect
- Officer Badge Number
- Officer Unit Number

Crime Incident Reports
- Date, Time, Location
- Crime Categorization
- Officer Badge Number
- Officer Unit Number

Officer Complaints
- Date and time
- Reason for Complaint
- Officer Badge Number
- Officer Unit Number

911 Responses
- Dispatch Category
- Date and Arrival Times
- Location
- Officer Badge Number
- Officer Unit Number

What would a National Repository of Policing Data enable?

It lets us to put the pieces together!

Parking Ticket
- Date, Time, Location
- Parking Violation
- Officer Badge Number
- Officer Unit Number

Municipal Citation
- Date and Time
- Municipal Offense
- Fine Amount
- Officer Badge Number
- Officer Unit Number

Investigatory Stop
- Date, Time, Location
- Reason for Stop
- Demographics of Subject
- Officer Badge Number
- Officer Unit Number

Traffic Stops
- Date, Time, Location
- Reason for Stop
- Demographics of Subject
- Officer Badge Number
- Officer Unit Number

Arrests
- Date, Time, Location
- Charges
- Demographics of Suspect
- Officer Badge Number
- Officer Unit Number

Crime Incident Reports
- Date, Time, Location
- Crime Categorization
- Officer Badge Number
- Officer Unit Number

Officer Complaints
- Date and Time
- Reason for Complaint
- Officer Badge Number
- Officer Unit Number

911 Responses
- Date and Arrival Times*
- Location
- Dispatch Category
- Officer Badge Number
- Officer Unit Number

Officer Activity Footprint and Police Unit Community Interaction Timeline

Individual officers, or their assigned units can be mapped over time, allowing a window into how they police communities and interact with victims.

What would a National Repository of Policing Data enable?

Participatory data governance standards

- **Maximal transparency given privacy concerns**

- **Generates (ethical) standards for research use**

Centralized access (and access controls)

- **Facilitate development and sharing of best practices**

- **Provide researchers a resource for secure and ethical data sharing**

Interoperability and comparative analysis

- **Maximal flexibility in how taxonomies are created and used for analysis**

- **Preserve taxonomic differences while enabling comparative analysis**

What would a National Repository of Policing Data require?

Direct engagement with community members

- Policing data are about community members, created using public funds

- Those policed have a normative (and, often, legal) right to responsible access

- Responsible access must be defined in concert with communities

Technical innovations

- Must preserve/track multiple taxonomies, sometimes within same system

- Must provide secure yet meaningful public access

Thought leadership

- What regulatory frameworks may be applicable? Why or why not? (**Future study!**)

- Public access often implies little legal oversight ➜ see engagement with community

Why now?

"Police are our government"

- **Responsible access is long overdue as a matter of government transparency**

- **Digitization of police administrative data makes this transparency feasible, at scale**

Sufficiently flexible database taxonomies

- **OWL-CHOIR ontology (see OCHRE database platform)**

- **Provides technical capacity to handle "messiness" of policing data, at scale**

New approaches to data governance

- **Participatory models of data governance (e.g., Pierre et al. 2021; Young et al. 2024)**

- **Secure third-party access is now possible, at scale**

Conclusion

The time is now, but the times are challenging!

A National Repository of Policing Data is feasible

- **Pilot use of OCHRE platform underway**

- **Significant amount of Chicago-specific contributed records**

Establishing this resource has major benefits

- **Laboratory for the study of privacy in (multi-modal) big data**

- **Enables previously impossible comparative analysis, without assuming a taxonomy**

Practical concerns about desirability

- **Participatory data governance is essential, but difficult**

- **Irresponsible access has significant ethical implications**

- **Current beliefs about policing in the United States are normative, often fear-based**

Acknowledgments

References

Comer, Benjamin P., Cody Jorgensen, and David Carter. 2023. "Reported Crime Frequencies: A Statistical Comparison of State Crime Reports and the UCR." *American Journal of Criminal Justice* 48(1):151–75. doi:10.1007/s12103-021-09623-y.

Soss, Joe, and Vesla Weaver. 2017. "Police Are Our Government: Politics, Political Science, and the Policing of Race–Class Subjugated Communities." *Annual Review of Political Science* 20 (1): 565–91. https://doi.org/10.1146/annurev-polisci-060415-093825.

Pierre, Jennifer, Roderic Crooks, Morgan Currie, Britt Paris, and Irene Pasquetto. 2021. "Getting Ourselves Together: Data-Centered Participatory Design Research & Epistemic Burden." In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 1–11. CHI '21. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445103.

Young, Meg, Upol Ehsan, Ranjit Singh, Emnet Tafesse, Michele Gilman, Christina Harrington, and Jacob Metcalf. 2024. "Participation versus Scale: Tensions in the Practical Demands on Participatory AI." First Monday, April. https://doi.org/20240428092301000.