

# Law Enforcement Management and Administrative Statistics (LEMAS) Datasheet

## I. MOTIVATION

### *I-A For what purpose was the dataset created?*

Conducted periodically since 1987, the LEMAS survey was designed to collect data on police agencies, including: agency responsibilities, operating expenditures, job functions of sworn and civilian employees, officer salaries and special pay, demographic characteristics of officers, weapons and armor policies, education and training requirements, computers and information systems, vehicles, special units, and community policing activities [1].

### *I-B Who created the dataset?*

Is it an official law enforcement or government body? An academic research team? Other?

The Bureau of Justice Statistics (BJS). The Relevant data expert at BJS is Elizabeth Davis.

### *I-C Was there a specific task in mind, or gap that needed to be filled?*

No extensive nationwide police agency survey existed before LEMAS.

## II. COMPOSITION

### *II-A What do the instances that comprise the dataset represent?*

For example: crimes, offenders, court cases, police officers

Each instance of the dataset is a survey response from a different police agency.

### *II-B Are there multiple types of instances?*

For example: offenders, victims, and the relationship between them.

No.

### *II-C How many instances are there in total?*

Of each type, if appropriate.

There are a total of 3,471 police agencies were sent the survey, including: 2,612 local police departments, 810 sheriffs' offices, and the 49 state agencies. A total of 2,779 agencies responded to the LEMAS survey, a response rate of 80%. The final dataset includes responses from 2,135 local police departments, 600 sheriffs' offices, and 49 state agencies.

### *II-D Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?*

For example, if it is traffic stops from a territory, is it all traffic stops conducted within that territory within a specific time? If not, is it a representative sample of all stops? Describe how representativeness was validated/verified. If it is not representative, please describe why.

The dataset includes all police agencies with 100 or more sworn personnel to be included, with smaller agencies sampled via stratified random sampling based on the number of sworn officers, and type of agency [2]. A total of 28 local police departments were determined to be out-of-scope for the survey because they were special jurisdiction agencies, had closed, had outsourced their operations, or were operating on a part-time basis [2].

### *II-E What data does each instance consist of?*

If there is a large number of variables, please provide a broad description of what is included.

Each instance consists of:

- Agency responsibilities.
- Agency expenditures.
- Officer salaries.
- Officer demographics.
- Agency policies.
- Officer requirements.
- Technology used at agency.
- Agency vehicles.
- Agency community policing practices.

### *II-F Is there a target label or associated with each instance?*

Please include labels that are likely to be used as target labels, e.g. recidivism.

No.

### *II-G Are there recommended data splits (e.g., training, development/validation, testing)?*

If so, please provide a description of these splits, explaining the rationale behind them.

No.

### *II-H Does the dataset contain data on race and ethnicity?*

If so, is it based on the individual's self-description, or based on officer's impression? Was it collected or derived in post-processing? For example, by name analysis.

Yes. The dataset contains information about the demographic composition of the officers within each agency. This

is (likely) based on self-description.

*II-I Are there any known errors, sources of noise, bias or missing data, or variables collected for only part of the datasets?*

If so, please provide a description.

No.

*II-J Does the dataset contain data on criminal history or other data that might be considered confidential or sensitive in any way?*

For example: sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; If so, please provide a description.

No. Demographics are not reported on an individual level.

*II-K Does the dataset contain data that might be considered confidential?*

For example: data that is protected by legal privilege or by doctor–patient confidentiality, data that includes the content of individuals’ nonpublic communications. If so, please provide a description.

No.

*II-L Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset?*

If so, please describe how.

No.

### III. USES

*III-A What type of tasks, if any, has the dataset been used for?*

If so, please provide examples and include citations.

The dataset has been used for a number of tasks, not limited to, but including:

- Investigating the effect of community policing practices [3], [4], [5].
- Investigating the changing demographics of police agencies [6], [7].
- Investigating technology being used by police agencies, including crime analysis tools [8], [9], [10].

*III-B Is there a repository that links to any or all papers or systems that use the dataset?*

If so, please provide a link or other access point.

Yes. Please see here:

<https://www.icpsr.umich.edu/web/NACJD/series/92/publication> earliest available data if from 1987.

*III-C What (other) tasks could the dataset be used for?*

For example: testing predictive policing systems, predicting recidivism.

The dataset could be used for any tasks which involve agency:

- Responsibilities.
- Expenditures.
- Salaries.
- Officer demographics.
- Policies.
- Officer requirements.
- Technology used.
- Vehicles.
- Community policing activities.

*III-D Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?*

For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other risks or harms (e.g., legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?

No.

### IV. COLLECTION PROCESS

*IV-A How was the data associated with each instance acquired?*

e.g. the data collected survey, the raw data is routinely collected by the courts.

The data was collected via survey.

*IV-B Was the information self-reported?*

If the data was self-reported, was the data validated/verified? If so, please describe how.

The information was reported directly from the agencies. However, it is organizations, not individuals, that were surveyed.

*IV-C Who was involved in the data collection process?*

Was this done as part of their other duties? If not, were they compensated?

Police agencies, no compensation was given.

*IV-D Over what timeframe was the data collected? Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)?*

If not, please describe the timeframe in which the data associated with the instances was created. If the collection was not continuous within the timeframe, please specify the intervals, for example, annually, every 4 years, irregularly.

The data is collected on an irregular basis. For example, the last collection was 2016, and before that was 2013. When collection takes place, it is done over the specified year. The

*IV-E Were any ethical review processes conducted (e.g., by an institutional review board)?*

If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

Unknown.

*IV-F Were the individuals in question notified about the data collection? Did they give their consent?*

If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?

N/A. The data is not on an individual level.

*IV-G If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?*

If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).

N/A

*IV-H Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted?*

If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

N/A

## V. PRE-PROCESSING, CLEANING, LABELING

*V-A Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, removal of instances, processing of missing values)?*

If so, please provide a description and reference to the documentation. If not, you may skip the remaining questions in this section.

The codebook does not specify pre-processing [11].

*V-B Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?*

If so, please provide a link or other access point to the “raw” data.

Skip.

*V-C Is the software that was used to preprocess/clean/label the data available?*

If so, please provide a link or other access point.

Skip.

## VI. DISTRIBUTION

*VI-A Is the data publicly available? How and where can it be accessed (e.g., website, GitHub)?*

Does the dataset have a digital object identifier (DOI)?

The dataset is distributed freely at:  
<https://www.icpsr.umich.edu/web/NACJD/series/92>

*VI-B Is the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?*

If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

The data-use statement says the following:

### Citation Requirement:

Publications based on ICPSR data collections should acknowledge those sources by means of bibliographic citations. To ensure that such source attributions are captured for social science bibliographic utilities, citations must appear in footnotes or in the reference section of publications.

### Deposit Requirement:

To provide funding agencies with essential information about use of archival resources and to facilitate the exchange of information about ICPSR participants’ research activities, users of ICPSR data are requested to send to ICPSR bibliographic citations for each completed manuscript or thesis abstract. Visit the ICPSR Web site for more information on submitting.

## VII. MAINTENANCE

*VII-A Who will be supporting/hosting/maintaining the dataset?*

The Bureau of Justice Statistics.

*VII-B How can the owner/curator/manager of the dataset be contacted (e.g., email address)?*

The BJS can be contacted at: [askbjs@usdoj.gov](mailto:askbjs@usdoj.gov)

*VII-C Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?*

Unknown.

*VII-D If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were the individuals in question told that their data would be retained for a fixed period of time and then deleted)?*

If so, please describe these limits and explain how they will be enforced.

The dataset does not relate to people.

*VII-E Are older versions of the dataset continue to be supported/hosted/maintained?*

Yes. Data from previous years continue to be hosted and are available for download.

*VII-F If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?*

If so, please provide a description.

No.

## REFERENCES

- [1] U. D. of Justice: Office of Justice Programs Bureau of Justice Statistics, "Law Enforcement Management and Administrative Statistics (LEMAS), 2016," 2016.
- [2] E. Davis, "Law Enforcement Management and Administrative Statistics (LEMAS) Methodology," 2016. [Online]. Available: <https://bjs.ojp.gov/data-collection/law-enforcement-management-and-administrative-statistics-lemas#methodology-0>
- [3] A. J. Diehr and J. T. McDaniel, "Lack of Community-Oriented Policing Practices Partially Mediates the Relationship between Racial Residential Segregation and "Black-on-Black" Homicide Rates," *Preventive Medicine*, vol. 112, pp. 179–184, 2018.
- [4] M. J. Hickman and B. Reaves, *Community Policing in Local Police Departments, 1997 and 1999*. US Department of Justice, Office of Justice Programs, Bureau of Justice, 2001.
- [5] M. A. Sozer and A. V. Merlo, "The Impact of Community Policing on Crime Rates: Does the Effect of Community Policing Differ in Large and Small Law Enforcement Agencies?" *Police Practice and Research*, vol. 14, no. 6, pp. 506–521, 2013.
- [6] D. A. Sklansky, "Not Your Father's Police Department: Making Sense of the New Demographics of Law Enforcement," *J. Crim. L. & Criminology*, vol. 96, p. 1209, 2005.
- [7] H. H. Yu and S. Viswanath, "Women in State Law Enforcement: An Exploratory Trend Analysis," *The American Review of Public Administration*, vol. 52, no. 4, pp. 268–279, 2022.
- [8] D. Weisburd and C. Lum, "The Diffusion of Computerized Crime Mapping in Policing: Linking Research and Practice," *Police Practice and Research*, vol. 6, no. 5, pp. 419–434, 2005.
- [9] S. Hyland, *Body-worn Cameras in Law Enforcement Agencies, 2016*. US Department of Justice, Office of Justice Programs, Bureau of Justice, 2018.
- [10] B. A. Reaves, "Local Police Departments, 2013: Equipment and Technology," *Washington, DC: Bureau of Justice Statistics*, 2015.
- [11] U. States Department of Justice. Office of Justice Programs. Bureau of Justice Statistics, "Law Enforcement Management and Administrative Statistics (LEMAS), 2016, Codebook," 2016. [Online]. Available: <https://www.icpsr.umich.edu/web/NACJD/studies/37323>