

National Incident-based Reporting System (NIBRS)

Datasheet

I. MOTIVATION

I-A For what purpose was the dataset created?

NIBRS was created to improve the overall quality of crime data collected by law enforcement. It aims to provide useful statistics to promote constructive discussion, measured planning, and informed policing. Giving context to specific crime problems such as drug/narcotics and sex offenses, as well as issues like animal cruelty, identity theft, and computer hacking. It intends to provide a nationwide view of crime based on the submission of crime information by law enforcement agencies throughout the country, offering law enforcement and the academic community more comprehensive data than ever before available for management, training, planning, and research [1].

I-B Who created the dataset?

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

NIBRS is collected and managed by the Federal Bureau of Investigation (FBI). Data is submitted by participating agencies.

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

NIBRS is an extensive dataset, collecting information on all *Group A* police incidents from across the United States. Including:

- Arson
- Assault Offenses
- Bribery
- Burglary
- Counterfeiting / Forgery
- Destruction of Property
- Embezzlement
- Fraud Offenses
- Gambling Offenses
- Homicide Offenses
- Human Trafficking
- Kidnapping / Abduction
- Larceny / Theft
- Prostitution Offenses
- Robbery
- Sex Offenses
- Weapon Law Violations

As such, it's potential uses are multi-faceted. It has not been created with a specific task in mind, but as a national centralized repository of police incident data.

The FBI has been reporting aggregated crime statistics through the *uniform crime reporting (UCR) summary reporting system (SRS)* since 1930. NIBRS is aimed at improving on UCR by reporting detailed information on an incident level, allowing for more detailed analysis.

II. COMPOSITION

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

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

In NIBRS instances are recorded crime incidents. An incident is defined as a set of offenses committed by one or a group of individuals, at the same time and place.

II-B Are there multiple types of instances?

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

Incidents are the 'base' unit of NIBRS. Each incident is linked to an agency and may be linked to one of more: offenses, offenders, victims, properties.

II-C How many instances are there in total?

Of each type, if appropriate.

In 2019, there were just under 7.7 million incidents recorded in NIBRS.

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.

NIBRS contain all incidents recorded by **participating** agencies. Incidents recorded by non-participating agencies are not included. Additionally, this is **not** a record of **all crime**. Only a subset of crimes are every encountered by police, and a subset of those are recorded as incidents.

NIBRS contains population coverage information, it can be determined how representative the incidents recorded are of the jurisdiction in which the agency operates.

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 instances contains the following information:

- Incident Information

- Incident Date
- Incident Hour
- Exceptional Clearance
- Exceptional Clearance Date
- Offense Information
 - Offense Codes
 - Attempted vs. Completed
 - Offender Suspected Use (of alcohol, drugs, or computers)
 - Location
 - Type and Number of Premises Entered
 - Type of Criminal Activity/Gang Information
 - Weapon/Force Used
 - Bias Motivation
- Property Information
 - Loss Type
 - Property Description
 - Value of Property
 - Date Recovered
 - Number of Motor Vehicles Stolen/Recovered
 - Drug Types and Amounts
- Victim Information
 - Connection to Offenses
 - Type of Victim
 - Age/Sex/Race/Ethnicity/Resident Status of Victim
 - Assault and Homicide Circumstances
 - Injury Types
 - Relationships to Offenders
- Offender Information
 - Age
 - Sex
 - Race
 - Ethnicity
- Arrestee Information
 - Arrest Date
 - Type of Arrest
 - Arrest Offense Code
 - Arrestee Weapons
 - Age/Sex/Race/Ethnicity/Resident
- Status of Arrestee
 - Disposition of Minor
 - Group B Arrest Information
 - Type of Arrest
 - Arrestee Weapons
 - Age/Sex/Race/Ethnicity/Resident
 - Disposition of Minor

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.

There is no set target label, though a few of interest may be: whether or not an arrest was made, the type of arrest, exceptional clearance.

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.

There is no official split. However, some points to consider:

When splitting data into multiple sets, be aware that the data is a single database that has been compiled from many agencies. If one wishes to test a predictive model, it may be reasonable to split along agency lines, assessing performance on unseen agencies.

If a temporal model is being used, to predict future offense numbers for example, the above is not applicable. Instead, it would make sense to have the same agencies across each split, with each split containing a different time segment.

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. Race and ethnicity are entered based on the officer's impression, in principle. In practice, it may be that in some instances the individual is asked about their race or ethnicity. These instances can not be distinguished. In addition, the ethnicity field is not used by all agencies.

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.

There are a number of fields which are officer estimates, and thus error prone: race, ethnicity, value of property, and drug amount.

In addition, value of property, and drug amount seems to sometimes be filled standardized amounts (1, 10, etc.). The policy regarding filling in those variables may differ between agencies.

II-J Is the dataset self-contained, or does it link to or otherwise rely on external resources?

For example: websites, tweets, other datasets)

The data is self-contained.

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.

The data contains records of crimes, some of which are violent. However, descriptions are minimal. Demographic information is recorded on both offender and victim. Additionally, it identifies whether the offense committed was a hate crime against any marginalised group, including LGBTQ+.

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 Has the dataset been used for any tasks already?

If so, please provide a description.

The dataset has been used in many studies. Including, but not limited to:

- Investigating the effect of demographics on incidents / arrests [2], [3], [4]
 - Investigating hate crimes [5], [6], [7].
 - Investigating crimes on juveniles [8], [9].
- among many others.

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.

The Inter-university Consortium for Political and Social Research (ICPSR) provide a non-exhaustive repository of publications using NIBRS data at:

<https://www.icpsr.umich.edu/web/ICPSR/series/128/publications>

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

This dataset can be used for investigating crime, where a significant amount of time, location and offense information is required. It is a highly flexible dataset that can answer many research questions when used correctly.

III-D Is there anything about the composition of the dataset or the way it was collected and pre-processed/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?

NIBRS is a collection of incident records, recorded and provided by thousands of police agencies. While NIBRS attempts to enforce standardisation, each agency will have it's own idiosyncrasies in recording. Some agencies do not record ethnicity, or use different units for recording drug quantities, among other differences. It is important to control for these differences when performing analysis on NIBRS.

Incidents that are related in real life cannot be connected within NIBRS. For example, a crime the occurred in the same time and place with two offenders who committed the same offense but one committed an additional offense will be recorded as separate incidents recording in NIBRS. There is no direct manner to connect these, so counting the same incident multiple times is possible if not careful. In addition, there are no unique identifiers for offenders or victims. Two offenses committed by the same offender at different times will not appear connected.

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.

Incident information is collected by and updated by each respective police agency using their own respective systems as the events occur. Once a year, incidents recorded by a participating agency are converted from their format to the NIBRS format, with help from the state UCR program. This data is then reported to NIBRS.

IV-B Was the information self-reported?

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

No. The data is recorded by police officers. However, some crimes may be recorded via victim's reporting.

The data is quality controlled and validated twice, once by state UCR programs, and again on reception by the NIBRS program.

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

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

Local police agencies. Data is recorded as part of routine police work.

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 has been continuous collected since 1988. However, the level of agency participation has changed during the years. For some states, data is available from 1998 onwards.

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?

Individuals may have known data is recorded. However, consent was not granted as the Individuals do not have the option to opt out.

IV-G 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.

Unknown.

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.

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

V-B Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data?

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

Police agencies will have local records that make the raw data sent to the NIBRS program, but these cannot be accessed.

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.

No.

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 available for download on the FBI's crime explorer website:

<https://crime-data-explorer.fr.cloud.gov/pages/downloads>

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 dataset is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) License.

VII. MAINTENANCE

VII-A Is the dataset maintained? Who is supporting/hosting/maintaining the dataset?

The FBI.

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

The owners can be contacted at: UCR-NIBRS@fbi.gov

VII-C Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (e.g., mailing list, GitHub)?

The dataset is published annually. Occasionally UCR will publish blocks of years, e.g. 2000-2010.

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

New data is released annually.

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

Yes. Data from previous years remains available for download from their website.

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] Federal Bureau of Investigation. National Incident-based Reporting System (NIBRS) User Manual. 2020.
- [2] Roland Chilton and John Jarvis. Using the National Incident-Based Reporting System (NIBRS) to Test Estimates of Arrestee and Offender Characteristics. *Journal of Quantitative Criminology*, 15(2):207–224, 1999.
- [3] Roland Chilton and Wendy C Regoeczi. Impact of Employment, Family Structure, and Income on NIBRS Offense, Victim, Offender, and Arrest Rates. *Justice Research and Policy*, 9(2):9–29, 2007.
- [4] Bradley Butcher, Chris Robinson, Miri Zilka, Riccardo Fogliato, Carolyn Ashurst, and Adrian Weller. Racial Disparities in the Enforcement of Marijuana Violations in the US. *AAAI/ACM Conference on AI, Ethics, and Society 2022*, 2022.
- [5] Kevin Strom. *Hate Crimes Reported in NIBRS, 1997-99*. US Department of Justice, Office of Justice Programs, Bureau of Justice . . . , 2001.
- [6] Michael Shively, Rajen Subramanian, Omri Drucker, Jared Edgerton, Jack McDevitt, Amy Farrell, and Janice Iwama. Understanding Trends in Hate Crimes Against Immigrants and Hispanic-Americans. 2014.

- [7] Yan Zhang, Lening Zhang, and Francis Benton. Hate Crimes Against Asian Americans. *American journal of criminal justice*, pages 1–21, 2021.
- [8] Melissa Hamilton. The biased algorithm: Evidence of disparate impact on Hispanics. *Am. Crim. L. Rev.*, 56:1553, 2019.
- [9] David Finkelhor. *Kidnaping of juveniles: Patterns from NIBRS*. US Department of Justice, Office of Justice Programs, Office of Juvenile . . . , 2000.