Improving Marylanders’ Safety & Prosperity Through Actionable Crime Data

Recommendations to the Maryland General Assembly

# Context and Vision

Since 1972, Maryland has published crime data summaries in line with the FBI's Uniform Crime Reporting (UCR) program. In 2021, the FBI directed states to transition to the National Incident-Based Reporting System (NIBRS), which has existed since the 1990s. Enrollment in NIBRS is at the law enforcement agency level and once an agency transitions, it ceases providing UCR data. Maryland agencies are currently in transition, with the 2021 UCR report separately reporting UCR and NIBRS data. Examples of this data are attached.

The UCR program has significant limitations. UCR provides summarizations rather than discrete crime data, and it only records the most severe offense in incidents with multiple offenses. This has resulted in an underestimation of crime rates by 10-14%. Additionally, offenses use a dated classification system for crime that engenders multiple flaws. Despite a significant allocation of resources to “Part 2” (lesser) offenses, UCR primarily summarizes “Part 1” (more severe) offenses. This leads to a further understatement of crime rates and prevents a full accounting of the volume of work police departments do and the resources they need to be effective.

NIBRS offers improvements such as providing data on every crime incident as well as a complete accounting of offenses in each incident. It provides a much more effective crime taxonomy and categorization, classifying offenses as crimes against people, property, or society, and reports on different crime categories equally. This makes it easier to shift crime intervention priorities based on data. Additionally, NIBRS provides metadata on victims, offenders, their relationships, property values, 911 calls, and more. However, NIBRS also has significant limitations preventing achievement of any outcome driven vision for criminal justice.

One limitation is lack of location data. Our data should enable us to identify and visualize emerging *local* crime patterns in real-time, along with their full economic and social context, to launch tailored interventions involving government, police, and community.

Many jurisdictions including Montgomery and Prince George’s already provide location data with every crime, but this is not collated at the state level. While staffing additional technical teams throughout state and local agencies to overhaul technology is advisable, it need not block immediate progress. Advances in technology allow people to create useful, peer-verifiable reports using commodity technology if we simply provide the data we already have. We can even fulfill some of the UCR program's vision to track each offender arrested in Maryland through the entire criminal justice system, from incident to arrest to courts to jails and prisons. Currently, crime and arrest data, trial data, and prison data are siloed among police departments, courts, and prisons, but this can be resolved with modest changes.

# Objectives and Recommendations

Suggested improvements to our crime reporting should rely on the following objectives:

* Location awareness: Instead of generalizing crime data to county or state levels leading to biased perceptions and reputations, we need precise, neighborhood-level crime data to understand where exactly different kinds of problems are.
* Transparency: to improve community safety, communities must build trust with law enforcement and align on shared outcomes. This requires data to be shared openly and accessibly to empower communities to develop deep knowledge of and participation in crime intervention.
* Actionability: Replace the broad, catch-all "crime rate" with targeted, local, crime-category-specific, contextualized, transparent metrics that drive public safety outcomes, enabling swift, focused responses to emerging patterns.

The recommendations herein are categorized by complexity and ease and are in the purview of the General Assembly to permanently mandate with the goal of reducing fluid political influences on and characterizations of crime and crime intervention. These would ideally be implemented in phases, with tier 1 passing immediately, and working towards tier 3.

## Tier 1: Immediate and straightforward actions

1. Require MSP publish all NIBRS data on Maryland OpenData and Maryland iMAP **including longitude and latitude for each offense (or incident)**.
2. Require Maryland Courts to enable a digital export of its *entire* digitized court records database for defendant name, docket number, charge, and relevant dates (what you see on search results page, not search details page). The existing web system requires specific search parameters for data access.
3. Ensure all Maryland police agencies transition to NIBRS by the end of 2023 to enable standardized delivery of statewide incident-based crime data by the end of 2024.

## Tier 2: Medium-term actions that may need additional support

1. Require Maryland Courts to include the referring police agency and the relevant offense/arrest record identifiers in its public court record, and require they make their entire database including all individual case records exportable.
2. Require MSP to include police officer badge numbers in all agency NIBRS reporting.
3. Require Maryland Courts to include prosecutor, judge, public defender, and defense attorney information and demographics in its exportable database, redacting personally identifiable information (PII) when necessary. This enhances legal system transparency, benefiting all parties and communities and enabling data-driven improvements.
4. Require MSP to publish the actual incident/offense/arrest report text alongside each NIBRS offense, redacting all victim PII.
5. Require Maryland Division of Corrections to publish a daily database containing relevant information for each incarcerated individual in state prisons AND jails. Records should include charge(s), incarceration date, projected release date, parole eligibility date, facility, age, race, gender, inmate identifier, and state court docket/case number. Names of incarcerated individuals need not be included (though it is recommended).
6. Allocate additional funds to MSP to hire records investigators, data scientists, and software engineers in the Central Records Division to improve data analysis, accuracy, and trend identification.

## Tier 3: Longer-term; high impact & ambitious

1. Address the challenge of disparate Records Management Systems (RMSs) used by different police agencies by requiring the Governor to select and contract a single RMS for all state law enforcement agencies. Fund the evaluation, acquisition, and transition costs to reduce crime reporting complexity, increase crime reporting reliability, consistency, and standards, and promote cross-agency collaboration.
2. Mandate that all state law enforcement agencies publish a digitized audit log of specific record changes from their RMS, including offense or arrest upgrades or downgrades after initial entry.
3. Initiate a comprehensive statewide effort for each agency to publish public data at the census tract level, utilizing Maryland OpenData and Maryland iMAP to publish a layered interface for public safety and related social and economic data in a single, detailed view. Most census tracts exist within county boundaries.
   1. Formalize the way the State Comptroller publishes State of Income (SOI) information, providing tax return metadata using percentiles for federal AGI at the census tract level. This will enhance income/wealth and wealth disparity reporting at the community level in a far superior way to the Census Bureau’s American Community Servey.
   2. Require the Department of Human Services to publish monthly summary application counts, active case counts, and metadata for welfare, unemployment insurance, disability, child support, child protection, and food/utility benefits at the census tract level.
   3. Require the Department of Education to publish GIS-compatible data showing the geographical boundaries serviced by each school, including associable "school report card" information and school demographics.
   4. Require and fund the Department of Budget and Management or another relevant state agency to produce monthly cost of living changes at the census tract level, along with an auto-generated annual report and statewide dashboard. This should include rents, home values/sales, utility costs, and costs of those goods used in the federal CPI basket, helping to contextualize affordability in relation to other social and economic trends.

# Author’s Background & Motivations

I am passionate about making my community and the world a better place. Over the last 15 years, I’ve channeled this passion at mission driven organizations including The New York Times, National Geographic, and the US Government Accountability Office. I started my career as a founding software engineer of DealerOn, a successful SaaS company in Rockville, Maryland where I was born and raised. I hold a BS in Computer Science from the University of Maryland, College Park.

I recently went on sabbatical with the intention of making a more direct impact on communities in Maryland. My vision was to build a tool that would allow any individual or agency to plot income, crime, employment, education, welfare, and public health data at the census tract level so that they could understand their communities better and drive more effective and targeted change. I first wrote a program that maps and colors each census tract in Maryland according to household income levels. The map was enlightening and at times surprising, expanding my understanding of our diverse communities. But when I started looking at crime data, I failed to find anything that went deeper than county-wide statistics.

How can policymakers, businesses, researchers, and the community itself address crime in context if they can’t place it on a map and identify trends? How can the state fund solutions if it does not have a statewide view of *where, exactly,* crime is happening?

This question ignited a weeks-long investigation into crime data publishing in Maryland. The more I learned about how crime is reported and the more I investigated the actual data, the more opportunities I found for improvement. I was particularly encouraged to see individual counties publishing their crime data including with location. My hope is that I can leverage my understanding of crime data as a data scientist and my background as an organizational leader to be a change agent for how we address crime in Maryland.

This report presents my findings and recommendations to the General Assembly, Governor, and Maryland State Police Superintendent and Central Records Division director, highlighting the importance of publishing reliable, transparent, local, and actionable crime data. By making such data accessible, I hope that collaboration and trust among citizens, law enforcement, and government agencies will bloom and lead to more safety and prosperity in every community.

# Methodology and findings

Work In Progress

High level statistical analysis of summary data provided in the PDF UCR report.

Forensic analysis of individual crime data in the Maryland NIBRS relational database published by the FBI. Analyzed using data queries using SQL programming language, not Excel sheets leveraging pivot tables and manual data cleansing. The following is a screenshot of some of the forensic level work done, which breaks down the offense count, arrest count, exceptional clearance count, and clearance rate for each type of crime in every county – far more detail than has been presented anywhere else.

A picture containing table

Description automatically generated

Difficulty in proving or disproving the summary NIBRS data in the UCR report because the database I’m using for forensic analysis from the FBI only includes agencies who have provided full year NIBRS data, but many have not, which means their data appears in the UCR report but not in the FBI database. This will require deeper analysis at the county level, but precludes verification of the clearance rate based on metadata because metadata is provided only at the state level and not at the county level in the UCR report.

In depth reading and analysis of UCR reports 2019-2021 (latest available). Research on transition from UCR to NIBRS. Analysis of summary data and identification of common weaknesses:

* UCR structure that is inherently flawed, such as breaking out “breaking and entering” separately from homicide and robbery. In many cases, breaking and entering is in furtherance of such a crime (known).
* UCR data that has been historically flawed, and those particular trends continue in NIBRS; detailed in the next few sections.
* NIBRS structure that is inherently flawed, for example allowing a single arrest to clear all offenses in an incident (known).
* Drug crimes are not separated by production/possession/intent to distribute (not elsewhere acknowledged).
* Summary NIBRS details data that are too sparse, vague, or implausible to justify the clearance rate (not elsewhere acknowledged). Most could just be recording process anomalies and does not necessarily imply manipulation.
  + NIBRS summary data in the UCR report does not break down fields such as relationship between offender and victim, offender ethnicity, or other metadata that would justify the clearance rate. In some cases, these numbers are so far off even in aggregate to justify the clearance rate. This requires forensic analysis
  + Homicide
    - Clearance rate is far lower in NIBRS than in UCR, which is not justified by the structural differences between the two.
    - Victim relationship field has an option called “not reported”, of which there are 125, almost half of the homicides, and greater than the clearance rate, casting doubt on the clearance rate. “Not reported” should be a very rare exception indicating a data flaw, as “Unknown” should be used for uncleared homicides.
    - 104 homicides listed as “unknown circumstance”, which seems high for the number of incidents/clearance rate.
    - A few cases where the victim relationship is listed as “victim was offender”, which is impossible for homicides.
  + Rape
    - TLDR: the record keeping here is really, really bad.
    - There are 276 clearances but only 193 arrests. It is normal and expected for arrests to be a little bit higher than clearances due to exceptional clearances or due to multiple people being arrested for a single offense, but this number is way too high. It implies that 83 cases occurred where a single arrest cleared multiple rapes.
    - 1413 rapes, almost half, are listed as “not reported” for the rape circumstance, which seems impossibly high and implies mass data collection issues.
    - 42% of offenders are listed as having unknown ethnicity, which should mean 42% of victims are unable to report the ethnicity of their rapist, which seems extremely high.
    - 2.8% of victims are listed as having unknown ethnicity, which seems like very poor record keeping.
    - 16% of offenders are listed as an unknown age, which doesn’t jive with the unknown ethnicity or the clearance rate.
    - Victim relationship field is troubling due to excessive options and poor guidance of what to choose. There are options for stranger, but also unknown and unreported. Either the victim knows who their offender is and we can pick a field (e.g. boyfriend), or the offender does not, and the offender is a stranger. There’s also an option for “family member” despite there are individual fields for each family member.
    - 5 cases where the relationship is listed as “victim was offender”, which is impossible except possibly for false allegation, but then that’s not rape and should not be counted as such.
    - The clearance rate of 18.6% is problematic. There is no clear definition for clearance of rape, as most rape victims know their assailants. It is not clear whether the low clearance rate is simply about not identifying the perpetrator or whether the police opted not to make arrests. The clearance rate of 20% for statutory rape makes no sense, as except for very rare cases of kidnapping, the minor will know the assailant and it should be highly exceptional for the police to not make an arrest in such a case, especially considering 889/1409 rapes are listed as having occurred in the home.
  + Robbery
    - Almost half are listed as occurring on “roadway”, which seems extremely improbably or very bad record keeping or both even when accounting for carjackings, evne if you suppose half of all robberies are carjackings. There are options for places like parking lot or gas station. It seems it would be exceptional for a robbery to happen on a “roadway” which implies vehicles are moving. This number may be including carjackings, and so something ‘on the road’ does not seem unbelievable. In short, carjacking reporting seems to be bad.
    - Same issues with weapon used as with homicide – most are not listed as knife or gun, which is completely implausible.
    - Same issues with victim relationship and ethnicity as rape and homicide
    - 15 listed as victim was offender, which is impossible.
    - Robbery property values are listed by type, but what would really be important to summarize is how many incidents happened per type of property, e.g. to say 20% of robberies were for cell phones, 30% for cars, etc.
  + Assault
    - Domestic violence cases are being undercounted, and are not being reported in line with changes Maryland made to its UCR reporting in the 90s which required DV to be broken out separately. Almost 8000/2400 (1/3) of incidents are listed as between boyfriend/girlfriend or ex-relationship, but the assault by circumstance field only lists 975 as domestic violence, and 19k/24k as circumstance not reported.
    - Table

      Description automatically generated with medium confidence
    - Same issues with weapon involved. Most list “personal weapon”, “other”, “unknown”, or “unreported”.
    - Relationship to victim field has issues as do the rest. 7752 cases list victim as offender, the most frequently used option, which is obviously nonsense.
  + Larceny
    - The vast majority is listed as shoplifting, from building, from motor vehicle, and motor vehicle parts. Yet 4500 incidents, more than half, are listed as having happened at a residence/home. No way.
  + Arson
    - 20 (10%) marked as other/unknown for location.
  + Race data for offenders across all Group A crimes is too sparse to justify the clearance rate, and race data for victims is also incredibly sparse even in cases of rape and homicide.
  + The “weapon used” field in both homicide and robbery has >40% use of “other weapon” or “personal weapon”, considering firearm/sharp object/knife/blunt object are options.
* Inconsistencies, anomalies, or impossibilities when deeply analyzing individual incidents in NIBRS data
  + 814 cases where arrests are listed without matching incidents (inconsistency)
  + Table

    Description automatically generated