Attachment 1 – Project Topic

Modeling the Criminal Case Process with State and Local Data (MCCP)

Key Objective

Multiple challenges impede integrating criminal justice data across state and local jurisdictions. Access to sensitive person-level or case-level data in any one jurisdiction has barriers. Data are dispersed and data systems have limited interoperability. Data about persons and cases within the criminal justice system are collected by multiple agencies—even in the same geographic area— and the agencies have different purposes for collecting data and different operational uses for it. Monitoring, assessing, and improving criminal justice policies and operations are strengthened by better understanding of the criminal case lifecycle. Improving data integration is necessary to produce more comprehensive statistics for evidence-building activities.

This Request for Solutions (RFS) seeks to test processes and tools that can be used to: (a) acquire and integrate operational data from multiple agencies within a given set of jurisdictions; (b) generate descriptive statistics and perform other analyses; and (c) document the solutions that can support a scalable and repeatable process for future data integration without requiring that agencies alter their data or systems.

Background

The National Secure Data Service

The CHIPS and Science Act, Section 13075(c), calls for engagement with federal and state agencies through an NSDS demonstration project to "collect, acquire, analyze, report, and disseminate statistical data in the United States and other nations to support governmentwide evidence-building activities consistent with the Foundations for Evidence-Based Policymaking Act of 2018."

Section 34 of the U.S. Code, Subtitle I, Chapter 101, Subchapter III established the Bureau of Justice Statistics (BJS) within the U.S. Department of Justice to collect and analyze "statistical information concerning crime, juvenile delinquency, and the operation of the criminal justice system" at the federal, state, and local levels. BJS is one of the principal statistical agencies in the U.S. federal government, an interconnected network of organizations that collects and transforms data into high quality statistical information, making it readily and equitably available to inform all types of decision-making, while protecting the responses of individual data providers or organizations.

The National Center for Science and Engineering Statistics (NCSES) at the U.S. National Science Foundation, like BJS, is a federal statistical agency with authority in law and expertise in practice to protect Americans' confidential data. The NSDS demonstration project at NCSES has clear statutory mandates about privacy protections, which call for NSDS to protect "confidential data and statistical products" and ensure that "no individual entity's data or information is revealed by the National Secure Data Service demonstration project platform to any other party in an identifiable form." Any datasets resulting from linkages with restricted data have the same protection as the restricted data. Linking restricted data with a public dataset, for example, results in a linked dataset that is itself restricted.

All NSDS projects using restricted data must be performing statistical activities for statistical purposes. These terms are defined in the <u>Confidential Information Protection and Statistical Efficiency Act</u> (CIPSEA) as follows:

"(10) STATISTICAL ACTIVITIES.—The term 'statistical activities'— "(A) means the collection, compilation, processing, or analysis of data for the purpose of describing or making estimates concerning the whole, or relevant groups or components within, the economy, society, or the natural environment; and "(B) includes the development of methods or resources that support those activities, such as measurement methods, models, statistical classifications, or sampling frames.

"(12) STATISTICAL PURPOSE.—The term 'statistical purpose'— "(A) means the description, estimation, or analysis of the characteristics of groups, without identifying the individuals or organizations that comprise such groups; and (B) includes the development, implementation, or maintenance of methods, technical or administrative procedures, or information resources that support the purposes described in subparagraph (A).

The Criminal Justice Process

State-level criminal justice systems in the United States function as distributed, multi-agency ecosystems spanning local, county, and state jurisdictions. A typical case lifecycle initiates with a law enforcement call-for-service, potentially resulting in an incident report. An arrest may then be made, followed by a decision on whether the person arrested will be detained pretrial or released on bond. Next, the applicable prosecutor's office will decide whether to file charges. If charges are filed, the case against the defendant will be adjudicated in court. This could involve an administrative response, such as a nolle prosequi or a plea agreement,¹ or it may involve a court trial. If the defendant is convicted (either through a plea agreement or a trial outcome), the case moves into the corrections phase, involving either community supervision or custodial confinement.

A model of the criminal case process has seven stages: arrest, prosecution, pretrial release, adjudication, sentencing, corrections, and appeals. These stages in the criminal case process are managed by distinct agencies that operate information systems that are mostly independent. Each system is built for domain-specific functions such as Computer Aided Dispatch (CAD) and Records Management System (RMS) platforms for calls-for-service and crime incident reporting, case management systems for prosecution and courts, or offender management systems for correctional supervision. These systems rely on disparate data models, schemas, and standards, and generally lack native interoperability, making end-to-end case tracking across agencies technologically challenging. Additionally, critical data elements may be embedded in unstructured formats (e.g., PDF reports, scanned documents, narrative text), further impeding data standardization, integration, and cross-system analytics.

¹ "Nolle prosequi" is a legal term used in both civil and criminal cases where the prosecutor decides to voluntarily drop the charges against a defendant before trial or before a verdict. See https://www.law.cornell.edu/wex/nol_pros for more information.

The U.S. Federal justice system is similar, in that data about criminal cases handled by Federal entities are maintained in separate data systems managed by different Federal agencies with limited, if any, interoperability. BJS has mapped the case processing for federal criminal justice data through its Federal Justice Statistics Program (FJSP). In this program, BJS links records from six agencies within the federal criminal justice system—US Marshals Service, Drug Enforcement Agency, Executive Office of US Attorneys, Administrative Office of the US Courts, US Sentencing Commission, and Bureau of Prisons. The linking process results in a data asset that allows BJS to describe patterns about the time from arrest through the conclusion of time under federal criminal justice control, whether that is a case declined for prosecution or release from post-incarceration supervision. Additionally, the linked data files allow for recidivism analysis related to federal offenses.

Record linkage is much more complicated at the state and local level, however, due to the large number of agencies involved, the proprietary nature of the data, data retention and access policies, and other factors. The fragmentation of records leads to an inability to discern patterns in criminal case processing and outcomes for state and local criminal violations, which make up the bulk of criminal justice in the United States. This RFS aims to lay the foundation for an integrated system for state offenses using data from state- and local-level criminal justice agencies.

A comprehensive data model that integrates person- or case-level data would allow stakeholders including governments, policy makers, practitioners, and researchers—to assess how effectively criminal justice systems address crime in their communities and to inform how criminal justice agencies can efficiently allocate time and resources to improve public safety outcomes.

Establishing a data model of how criminal cases are processed at the state and local level will provide answers to important justice and public safety questions such as –

- 1. How long do cases take to be processed from start to finish?
- 2. What is the case management time associated with each stage of the process?
- 3. How often do arrests result in prosecution?
- 4. How often do prosecutions result in a conviction?
- 5. How often are convictions the result of plea bargains or of a trial?
- 6. Which types of sentences are given and how long are they?
- 7. How often do offenders recidivate?

Additionally, these outcomes could be compared across jurisdictions, offense types, demographic characteristics of the victims or offenders, and other factors.

The Current Request for Solutions

The Modeling the Criminal Case Process (MCCP) project aims to evaluate tools and methodologies for:

- developing a privacy-preserving record linkage framework that enables secure integration of case- or person-level data from multiple state and local sources across the criminal justice ecosystem while maintaining confidentiality;
- 2. applying artificial intelligence techniques to process both structured and semi-structured case data; and
- 3. delivering support services to data providers, including subject matter expertise and technical assistance.

At least two agencies from three jurisdictions, including but not limited to law enforcement agencies, prosecutor's offices, courts, jails, community corrections organizations, and prison systems, will be recruited to provide access to record-level data for criminal incidents, arrests, prosecutions, court filings, court case outcomes, and custody admissions and releases. In a given jurisdiction, the two agencies should be as close in proximity in the stages of the criminal case process so that a descriptive statistic on the proportion of cases that convert or last from one stage to the next successive stage can be calculated. Additionally, BJS federal data support that linkages between agencies close in the process have better and more stable link-rates than agencies further apart in the process. Record-level arrest and prosecution data should be prioritized for inclusion.² Subject matter experts at the BJS plan to work with the awardee to identify state and local agencies that are likely to participate in the project if contacted. Records from the various agencies will be linked using privacy-preserving techniques. Additionally, this project seeks to explore the use of AI techniques to process both structured and semi-structured source data.

To work effectively with the data providers, the team will need to include one or more members with experience and expertise working with criminal justice agencies and requesting and using their data. For privacy-preserving methods to be effective, data providers need ample support from both subject matter experts and technical experts to prepare and share the data. The project will need to incorporate measures for helping agencies provide access to their data efficiently and effectively. The team will evaluate lessons learned from the assistance provided, which will help inform NSDS's planned concierge services. The team will produce a set of descriptive statistics and other analyses from the privacy-preserved data and provide an assessment of the data quality.

Information Gaps

If such a system were adopted and implemented -

• State and local governments, researchers, and others would benefit from knowing how individuals and cases progress through the criminal justice process. Better understanding about the entirety of the process, how long it takes, and how the process varies would be possible.

Key Evidence-Building Considerations

Modeling the Criminal Case Process (MCCP) will fulfill the following objectives:

- Assess the process used to acquire record-level data from relevant agencies, and identify possible improvements
- Assess the technical and organizational processes used to link the data, assess the resulting data quality, and identify possible improvements
- Develop mechanisms and tools for acquiring and linking operational case data across data collecting agencies in a privacy-preserving manner

² While the Federal Bureau of Investigation collects data on arrests made by local law enforcement agencies through the National Incident-Based Reporting System (NIBRS), that dataset alone will not provide the required information to establish linkages to other data sources.

- Leverage current best-practice frameworks and ongoing research to develop methodology for evaluating linkage quality, including an assessment of its applicability to other agencies and jurisdictions
- Explore the use of artificial intelligence to process and analyze semi-structured case records in a reliable and scalable way
- Assess the types of technical assistance needed to support local and state governments to provide data for future data integration efforts.

Deliverables

The following minimum deliverables will be provided:

- Project Management
 - Business Requirements Documentation outlining the broader goals and objectives of the MCCP.
 - Monthly status reports on progress towards project objectives.
 - Quarterly lessons learned based on what has been learned during the last quarter that will inform a future NSDS.
 - A final report on the project and its outcomes.
- Technical Documentation
 - Information architecture associated with the acquisition and linking of data sources, to include any additional capabilities required to process the data, such as server requirements, data storage, and security protocols.
 - Guidelines for acquiring and linking operational data for statistical and research purposes in a privacy-preserving manner.
 - Specifications for and codebase used in the processing and analysis of the data, sufficient to allow for reproducibility.
 - An assessment of the processes developed and the resulting data quality.
 - An assessment of the services provided to data providers.
- Statistical Analysis
 - o Descriptive statistics and other analyses conducted on dataset
 - All code used in the project provided in a reproducible, and if applicable, open-source format (R, python, etc.)
 - Data (and/or documentation of data) sufficient to reproduce the results of all statistical analysis including those done using open-source software
 - A report describing the key findings, the statistical methodology employed, and an assessment of the statistical quality.