WASHINGTON STATE STATISTICAL ANALYSIS CENTER

Criminal Justice Research & Statistics Center

Informing a data-driven justice system

The Justice Data Warehouse (JDW): The Data Handbook

Vasiliki Georgoulas-Sherry, Ph.D. & Hanna Hernandez, M.A.

Contents

About the CJRSC – the Washington SAC

SACs are found across the nation in 51 states and territories. As the Washington SAC, the CJRSC is responsible for collecting, analyzing, and reporting public safety and criminal justice related statistics to federal, state, and local levels of government. We also facilitate the sharing of state-level information nationally. The information produced by SACs and their involvement in criminal justice projects is critical to local, state, and federal criminal justice agencies and community organizations as they develop programs and policies related to crime, illegal drugs, victim services, and the administration of justice.

SACs play a vital role in developing criminal and juvenile justice policy at the state and local levels. Their research provides evidence that policymakers use to guide their decision-making. By furthering the use of evidence-based practices in their states, SACs promote the effective and efficient administration of criminal and juvenile justice.

Contact Us

Phone	Fax	Address	Website
360-902-0599	360-586-1988	P.O. Box 43113 Olympia, WA 98504-3113	sac.ofm.wa.gov

Acknowledgements

This CJRSC – the Washington SAC would not be possible without the help and support of the Department of Justice (DOJ)'s Bureau of Justice Statistics (BJS)' State Justice Statistics (SJS) grant.

Version History

Date	Version	Author(s)	Revision Note
09/03/2025	1.0	Vasiliki Georgoulas-Sherry, PhD	Document creation
02/18/2026	2.0	Vasiliki Georgoulas-Sherry, PhD	Document updated

Abstract

Washington state's criminal justice system has long operated in disconnected silos across federal, state, local, and Tribal levels, limiting the ability to assess performance, address disparities, and improve outcomes.

To respond to these impacts, the Washington Statistical Analysis Center (SAC) applied for and received the 2021 State Justice Statistics (SJS) grant from the Bureau of Justice Statistics (BJS). To tackle these challenges, the SAC, in partnership with the Public Safety Policy and Research Center (PSPRC), established the Justice Data Warehouse (JDW). This integrated platform links data from courts, jails, prisons, community supervision, and more, offering a comprehensive, longitudinal view of individuals' justice system involvement.

This document describes the JDW data at a high-level. The JDW can enhance transparency, supports data-driven policy, and enables cross-sector collaboration. By centralizing justice data and supporting local jurisdictions, the JDW can play a pivotal role in building a smarter, fairer, and more accountable justice system for Washington.

Background

Like other criminal justice and legal systems in the nation, Washington state is insular; each decision point of the criminal justice and legal system is sectioned into silos, which includes state and local agencies. Efforts continue to be made to shift from this siloed approach to a more integrated system.

Similar to other states, Washington lacks data integration capabilities that would allow for a comprehensive assessment of the criminal justice and legal systems' performance and outcomes. This is true at the local levels, as well as the state level. For example, measuring performance and outcomes across systems (such as jails, courts and juvenile detention, community supervision, and prison incarceration) are hindered by a lack of integration. Addressing the need for integration and promoting collaboration across different components of the criminal justice and legal system can help bridge these silos and create a more effective system.

With this in mind, the SAC, in conjunction with the PSPRC, created the JDW to integrate criminal justice databases together as currently, each agency maintains their own distinct data sets. The JDW represents a significant step toward centralizing disparate data systems and enabling a holistic view of Washington's justice system. By linking data across agencies and jurisdictions, the warehouse supports a unified approach to understanding how individuals move through the system and where interventions may be most needed. For example, the ability to track an individual's journey from arrest through court proceedings, incarceration, and community supervision provides insight into system bottlenecks, disparities in treatment or outcomes, and the effectiveness of rehabilitative programs. This capacity for longitudinal analysis is critical for developing policies rooted in evidence rather than assumptions.

Moreover, the JDW can strengthen transparency and accountability. Without integrated data, it is nearly impossible for policymakers, researchers, and the public to access meaningful information about systemic issues such as racial disparities and recidivism. With standardized and accessible data, interested parties can assess the impact of laws and policies in real time, uncover inequities, and identify areas in need of reform. The JDW can enable interested parties to see how individuals interact with multiple components of the justice system — from arrest to reentry — creating a fuller picture.

Using Integrated Justice Data

Using integrated administrative data from criminal justice agencies can harness information in meaningful ways that transcend traditional "silos" and allows communities to focus collective attention on important social issues that cross systemic boundaries. As individuals in the criminal justice system frequently interact with other decision points (e.g., policing/arrests, trial/sentencing, and incarceration/community supervision), collecting and analyzing data across multiple criminal justice systems is critical to better understand the impact of the system.

The JDW's integration would also allow us to pinpoint data to particular areas of intervention (e.g., overrepresentation of person of color at all stages of the criminal/legal system, disparities in outcomes for marginalized individuals who interact with the criminal legal system, etc.). Additionally, the JDW could also be used to track and improve fairness in areas such as prosecution, adjudication, disposition, sentencing, incarceration, release, and community supervision.

An integrated justice data system like the JDW is essential for addressing the limitations of fragmented data that hinder effective system-wide understanding and response. Without integration, agencies often operate in isolation, making it difficult to track an individual's full trajectory through the criminal justice system or assess the cumulative impact of decisions made at different stages. Integrated data allows interested parties to break down these silos and view the system as an interconnected whole. This holistic perspective is critical for identifying gaps in service delivery, or unintended consequences of policy decisions. It also ensures that reforms in one part of the system — such as changes in sentencing guidelines — can be evaluated in terms of their ripple effects throughout the rest of the system.

The benefits of integrated justice data extend beyond operational efficiency and into improved outcomes for individuals and communities. By enabling comprehensive analysis, the JDW can help illuminate patterns and trends that may otherwise remain hidden — such as racial and geographic disparities, repeat justice involvement, or systemic delays in case processing. These insights support data-informed decision-making that can lead to more effective, equitable, and transparent justice policies. Ultimately, a centralized data system enhances accountability, supports smarter resource allocation, and strengthens the state's ability to deliver fairer, more responsive justice to all Washingtonians.

Cross-Sector Research and Analysis in Justice Data

By fostering collaboration among criminal justice agencies, the JDW could ensure that justice system data is not only comprehensive but also actionable. A well-integrated repository allows interested parties to analyze trends across law enforcement, courts, corrections, and community supervision, helping to identify systemic gaps, high-risk populations, and opportunities for reform.

The integration of diverse data sources — including arrest records, court filings, jail and prison data, and recidivism rates — provides a more complete picture of the justice system's functionality and impact. The ability to track individuals' experiences across the justice continuum over time supports the identification of disparities, ensuring that vulnerable or marginalized groups receive the attention and resources needed to achieve fairer and more just outcomes. The JDW could serve as a critical tool for guiding evidence-based policy decisions that center on justice reform, equity, and effectiveness.

By leveraging cross-sector collaboration, the JDW will also support the development of proactive strategies that prevent justice involvement and reduce recidivism. With access to integrated data, researchers and policymakers can identify patterns and emerging risks, enabling early interventions. Ultimately, the JDW's comprehensive approach to justice data analysis will play a vital role in shaping policies that reduce harm, improve outcomes, and enhance public trust in the justice system.

Technological advancements and innovative justice policies can also benefit significantly from cross-sector collaboration. Research that integrates data from courts and law enforcement can highlight the impact of pretrial detention, sentencing practices, or reentry barriers on long-term outcomes. By analyzing this information, policymakers can implement targeted reforms, community-based supports, and policies that reduce incarceration while improving public safety and social equity. Cross-sector analyses also inform long-term strategies that promote restorative justice, reduce over-reliance on incarceration, and strengthen community resilience.

JDW Core Data

The SAC, with the support of the PSPRC, receives a variety of administrative datasets from agency partners that are incorporated into the JDW. These administrative datasets are outlined in Table 1, based on the category of data and data source – Washington State Administrative Office of the Courts (AOC), Washington State Department of Corrections (DOC), Caseload Forecast Council (CFC), Washington Association of Sheriffs & Police Chiefs (WASPC)'s Jail Booking Reporting System (JBRS) and Washington State Patrol (WSP)'s Computer Criminal History (CCH) data.

Table 1: The JDW Core Data Sources

Type of Data	Agency	Description
Carceral and Community Supervision	Department of Corrections (DOC)	DOC maintains information for people incarcerated in Washington state correctional facilities and for people under community supervision in Washington state.
Court Case Filings – Judicial Information System	Administrative Office of the Courts (AOC)	AOC maintains statewide electronic court records database for all cases seen by courts in Washington state.
CFC Convictions	Caseload Forecast Council (CFC)	CFC maintains the adult conviction database including data related to the crime, the offender, the sentencing judge, the sentence, and alternatives to incarceration.
Jail Booking and Reporting System (JBRS)	Washington Association of Sheriffs & Police Chiefs (WASPC)	JBRS (RCW 36.28A.040) is a multi-jurisdictional database providing criminal justice agencies an instant, up-to-date database of booking and release records from all city and county jails in Washington state.
Computerized Criminal History (CCH)	Washington State Patrol (WSP)	The WSP maintains a database of Washington criminal history information, or background checks, consisting of fingerprint-based records and disposition information from law enforcement agencies and courts throughout the state.

JDW Administrative Data Limitations

While all the datasets above are processed to the highest quality standards by the data contributing agencies, it is important to recognize that inaccuracies may exist within administrative data. Unlike other data, where both cross- and within-subject controls are possible, such measures are often unfeasible and impossible to incorporate in administrative data. Administrative data is also not typically collected for research or evaluation purposes but to meet the administrative needs of specific programs and specific state or federal reporting or monitoring requirements. Administrative data is collected as both transactional and summative datasets by local administrators and submitted to an agency authority, making variance among data collectors a potential source of bias in each dataset. Quality control processes may be imposed after data is submitted to agency authorities, which could impact data quality in ways that are difficult to detect within the final dataset. The limitations are not meant to suggest that the administrative data loaded into the JDW is unreliable but rather to advise individuals to keep these potential concerns in mind as they request data and conduct research. Administrative data must always be thought of as the combination of both the collected data and the process used to collect the data.

System Organization

The system comprising the JDW is organized around the following components:

(1) data loading, (2) data profiling, (3) data modeling, (4) data mappings and validation routines, (5) data linking and identity resolution.

Data Loading

Source data for the JDW repository is securely transmitted from providers using encrypted standard formats and loaded into prestaging and staging databases, where automated validation, conformity checks, and data quality flagging occur. Identity resolution variables are used for cross-sector linkage, after which linked data is transferred to the Operational Data Store (ODS). Personally identifiable information is excluded from the ODS and replaced with unique anonymized IDs to support linkage while protecting privacy. Research-specific analytical data files are then extracted from the ODS to support targeted analyses.

Data Profiling

With each data update, raw data is profiled by the SAC and PSPRC to confirm data validity and accuracy. When available, data dictionaries are used in the data profiling process, as are all the codes and lookups, business logic around computed variables, and programmatic information as it relates to data elements. Values, ranges, and frequency distributions are compiled and shared back with the data provider for confirmation or correction.

Data Modeling

When new or modified data sources are introduced, data modeling is performed to integrate related data into a unified structure for efficient storage and management. In integrated criminal justice records systems, relationships among entities such as people, crashes, and addresses inform the model design. Data modeling is also used to create subject-specific analytical files that support research and analysis.

Data Mappings and Validation Routines

When new or modified data sources are introduced, all data mappings, transformations, and validation rules are defined and documented for each dataset and loading step. Documentation includes data type changes, code standardization, and any transformation or categorization logic. Validation routines standardize formatting, verify completeness and referential integrity, validate code values, and generate data quality flags to identify invalid or suspect data for downstream processes and end users.

Data Linking and Identity Resolution

Each data source is evaluated to identify the optimal combination of data elements (“tokens”) for uniquely identifying individuals. Identity resolution variables are assessed for completeness, consistency, and quality over time to inform linkage rules. Deterministic and probabilistic matching methods are used to link records across datasets, with manual review conducted for uncertain matches. The process aims to maximize accurate person-level linkage while minimizing false matches.

Identity Resolution Process

The core feature of the JDW is the linking of cross-sector data. Through an identity resolution process, JDW links individuals across data files from contributing agencies to facilitate longitudinal and cross-sector analysis. Identity resolution is the process of identifying records that belong to the same entity (e.g. person or household). The purpose of JDW’s identity resolution process is to identify and create linkages across multiple data sources so that crash records associated with a given individual are linked to related records associated with that individual and event in the roadway, police interaction, court interaction, and health encounters datasets. For the JDW, this involves linking individual-level data, such as names and birth dates, across multiple sources and identifying these individuals with unique person identifiers. These identifiers are referred to as “PersonIDs.” PersonIDs are assigned

to all individual-level data received by SAC/PSPRC from our data contributors. As additional linking activities occur, PersonIDs are updated to reflect the most recent data available.

Creation of PKeys and Assignment of PersonIDs

Before individual-level PersonIDs can be created or assigned, identity resolution tokens referred to as a "PKeys" are created for each record in a dataset. A "PKey" is an identifier or combination of identifiers from a dataset that are unique to an individual within the dataset. JDW reviews each dataset to establish PKeys specific to the dataset. Identifier and individual characteristic data from the datasets and associated with the PKeys are then loaded into the identity resolution system for linkage. PKeys that already exist in the identity resolution system are attributed with the PersonID assigned to that PKey. PKeys that do not already exist in the identity resolution system are assigned a unique preliminary PersonID which may be overwritten in the identity resolution steps.

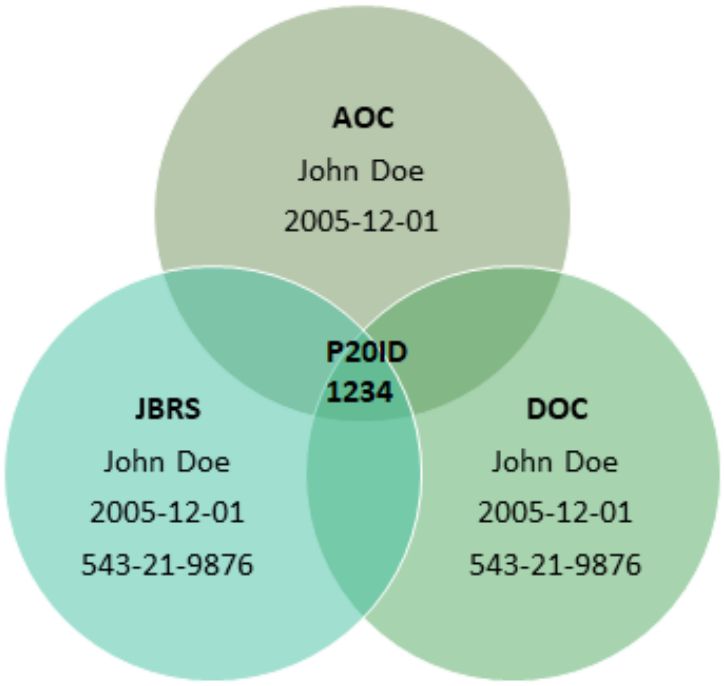
Because no common set of identifiers exists across all data sources, identity resolution processes and match rules are tailored to each dataset and limited to the identifiers available for linkage (see Table 2).

Table 2: Identity Token Components in Data Sources

Sector	Agency	Birthdate	First Name	Middle Name	Last Name	SSN
AOC	AOC	**X**	**x**	**x**	**x**	
CFC	CFC	**X**	**x**	**x**	**x**	
DOC	DOC	**X**	**x**	**x**	**x**	**x**
WASPC	JBRS	**X**	**x**	**x**	**x**	**x**
WSP	CCH	**X**	**x**	**x**	**x**	**x**

Two datasets with different sets of available identifiers can be indirectly matched by involving other sources of individual data. For example, the Licensing dataset above can only be directly linked to the Collision data, but the Collision data can be linked to all other datasets. Consequently, the licensing data may be linked *indirectly* to all datasets in the JDW repository (see Figure 2).

Figure 2: Identity Token Components in Data Sources



Phases of Identity Resolution

JDW's identity resolution process has four phases which are performed sequentially for each iteration of linkage.

Blocking

Blocking is the process of generating potential match pairs between new PersonIDs and existing PersonIDs within the identity resolution system using predefined match rules. These rules compare combinations of identifiers—such as names, birth dates, and SSNs—to identify records that may belong to the same individual. Match quality depends on the strength of the rule applied: exact matches across multiple identifiers generally produce low false positive rates, while rules based on fewer fields, partial strings, or fuzzy matching techniques increase the likelihood of uncertainty. After potential match pairs are generated, the identity resolution system applies the Expectation Maximization (EM) algorithm to calculate a probabilistic match score for each pair.

Evaluation

The set of potential match pairs is split into three categories:

- **High-probability matches** result from conservative match rules, such as exact agreement on name, birth date, and Social Security number. These rules are designed to produce extremely low false positive rates, making undermatching a minimal concern. Probabilistic match scores may also be used to further refine this set of highly reliable matches.
- **Mid-probability matches** are generated using less restrictive rules, such as matching on name and birth date only. Associated identifiers and probabilistic scores are compiled into a review dataset and manually evaluated to determine whether records represent true matches. Confirmed matches are flagged and incorporated into the identity resolution system as provisional match pairs.
- **Low-probability matches** are produced using broad or weak match criteria, such as shared county of residence, gender, and first name. Because these pairs typically receive low probabilistic match scores and rarely represent valid matches, they are generally excluded from further review and ignored.

Cardinality Analysis

Cardinality analysis is a key step in identity resolution that enables more aggressive matching while improving the quality of existing PersonID linkages. In this process, provisional match pairs are temporarily merged and evaluated to determine the relationships between PersonIDs in the subject dataset and those in the broader repository. These relationships may be 1:1, 1:Many, Many:1, or Many:Many—for example, a 1:Many relationship indicates one PersonID matches multiple PersonIDs in the dataset. 1:1 relationships are accepted, while non-1:1 relationships are manually reviewed to resolve discrepancies as accurately as possible. This step also allows analysts to reassess prior matches using new information, potentially merging or unmerging records from earlier iterations. As a result, identity resolution linkages continuously improve over time. Once verified, results are fed back into the system, where PersonIDs may be merged, unmerged, or provisional matches rejected.

Merging

After the cardinality phase is concluded, the match table now contains a list of positive match pairs of PersonIDs. These match pairs are then incorporated into the identity resolution system using an automated process. The result is that people who had been previously represented by multiple preliminary PersonIDs are now represented by a single PersonID.

Core JDW Contributor Data

This section provides a set of descriptions to the core data files that feed into the JDW. Please note, all data comes with its limitations. This report identifies a few major limitations, and there are likely more not listed that could impact work that utilize this data. This information is not an exhaustive list of data in the JDW.

Washington State Patrol

The Washington State Patrol (WSP) Computerized Criminal History (CCH) system is the state's central repository for criminal history record information. The CCH contains detailed information on individuals who have been arrested and fingerprinted in Washington. This includes data on arrests, charges, court dispositions, sentencing outcomes, and correctional supervision. CCH data is used extensively by law enforcement, the courts, licensing agencies, and employers for purposes ranging from criminal investigations to firearm purchases, child welfare screening, and employment eligibility.

As all data, the CCH system has limitations that can affect both accuracy and usability. Delays in data entry, inconsistent reporting practices, and technological disparities across local jurisdictions contribute to data fragmentation. These limitations highlight the need for improved data-sharing agreements, automation, and uniform reporting practices across the justice system.

Washington Association of Sheriffs and Police Chiefs

The Washington Association of Sheriffs and Police Chiefs (WASPC) Jail Booking Reporting System (JBRS) is a statewide effort that collects standardized booking information on individuals entering jails across Washington, including booking dates and times, charges, demographics, arresting agency, bail or bond amounts, and release information. It provides a snapshot of jail activity and the circumstances of detention. JBRS aggregates data through interfaces with locally managed jail record systems under RCW 70.48.100. Because jails are operated by counties, cities, and tribal nations, data completeness and structure vary by jurisdiction, and JBRS outputs represent point-in-time snapshots rather than a comprehensive statewide dataset. The system is focused on bookings only and does not include full arrest histories, complete release data for all individuals, or information from non-participating or non-overnight facilities, including some tribal jails and San Juan County Jail.

Reliance on external systems means JBRS is subject to human error, system variability, and interface outages, which can delay, omit, or mis-time stamp records. As a result, JBRS cannot guarantee the accuracy, completeness, timeliness, or fitness of the data and disclaims responsibility for any resulting errors or losses.

Caseload Forecast Council

The Caseload Forecast Council's (CFC) Judgment and Sentence (J&S) data is a crucial dataset that captures detailed information on felony convictions in Washington State, based on standardized sentencing documents submitted by Superior Courts. It includes offense type and severity, criminal history scores, sentence length, sentencing enhancements, and whether sentences are concurrent or consecutive, providing a key record of how sentencing laws are applied in practice.

However, the J&S data has limitations. Submissions may be delayed or inconsistent, and records may be missing key details. It also reflects sentencing at conviction only and does not capture later changes such as appeals, sentence modifications, or supervision updates. In addition, limited shared identifiers make it difficult to link J&S records with other justice datasets, constraining full system-wide analysis.

Washington State Administrative Office of the Courts

The Washington State Administrative Office of the Courts (AOC) serves as the central repository for data collected from courts across the state, including superior, district, municipal, and juvenile courts. The AOC manages and maintains comprehensive records related to court case filings, dispositions, sentencing, bail and bond decisions, warrants, and protection orders. These datasets are essential for understanding how cases are processed through the legal system and for analyzing outcomes at various decision points.

Like all data sets, there are limitations that exist such as data standardization and completeness across courts and jurisdictions. Because Washington's courts operate independently under the state's decentralized judicial system, data quality and reporting practices can vary widely. Not all courts use the same case management systems, and some may enter data inconsistently or omit critical fields.

Washington State Department of Corrections

The Washington State Department of Corrections (DOC) maintains extensive datasets on individuals under its supervision, including those incarcerated in state prisons and those on community supervision (formerly known as parole or probation). This data includes demographic information, conviction details, sentencing length, facility location, custody level, program participation, disciplinary infractions, and release information.

Discussion

The JDW represents a transformative shift in Washington State's approach to criminal justice data by moving from fragmented, siloed systems to an integrated, statewide data environment. Historically, criminal justice data has been distributed across federal, state, local, and Tribal agencies with limited interoperability, constraining the ability to assess system-wide outcomes, track individual trajectories, or identify disparities. The JDW addresses these limitations by centralizing data across the justice continuum, including arrest, adjudication, incarceration, and community supervision.

This integrated structure enables longitudinal analysis of system involvement, improves the ability to identify inefficiencies and inequities, and supports greater transparency and accountability through standardized, accessible data. It also strengthens the capacity for research and policy evaluation by allowing more rigorous analysis of sentencing patterns, system performance, and the impacts of legislative and programmatic changes. From a policy perspective, the JDW supports evidence-based decision-making and helps inform reforms aimed at improving system effectiveness and fairness.

Technological integration within the JDW further enables improved reporting tools, dashboards, and coordinated system monitoring, helping agencies better understand how decisions at one stage of the justice process influence outcomes across the system. Collectively, the JDW advances a more unified, data-driven approach to criminal justice in Washington, strengthening system understanding, improving accountability, and supporting more informed reform efforts.