CRIMINAL JUSTICE RESEARCH INSTITUTE

HAWAI'I

A REPORT ON THE CREATION OF A CENTRALIZED PRETRIAL DATA REPORTING AND COLLECTION SYSTEM, PURSUANT TO ACT 147, SLH 2023

PREPARED BY

Erin Harbinson, PhD and Aerielle Reynolds, MSCJA

BACKGROUND

This report is respectfully prepared pursuant to Act 147, Session Laws of Hawai`i 2023, which appropriated funds to the Criminal Justice Research Institute (CJRI) to create a pretrial database and reporting system, and required a progress report for the Legislature over the next two years. The appropriation from Act 147 provides CJRI with the funds to fulfill a mandate outlined in Act 179 (Session Laws of Hawai`i 2019), which requires CJRI to create a "centralized statewide criminal pretrial data reporting and collection system" (HRS § 614-3). This report is the first report of two that must be submitted, which summarizes progress made approximately six months since Act 147 was signed into law.

As documented by the Criminal Pretrial Task Force in 2019 [1], a major barrier to understanding how the pretrial system is operating comes from the siloed and disconnected data across agencies involved in the pretrial system. The Criminal Pretrial Task Force recommended that a Criminal Justice Research Institute be created in order to bring data across these agencies into a centralized system. This system would create capacity to report out on criminal pretrial metrics. To accomplish this mandate, the law recognized that CJRI must take several steps to develop a plan and solution to create the database. These steps acknowledged in the law include: 1) identifying databases with pretrial information, 2) determining the administrative and technological feasibility of aggregating and sharing current data, and 3) identifying gaps in pretrial data (HRS § 614-3). CJRI staff completed these steps and identified several challenges that must be addressed in the project:

- Currently, there is some data, but much of it is still stored as information and therefore information must be transformed into data for statistical analysis.

- Two branches of government house pretrial data, which means different laws and rules govern their data use, as well as differences in administrative practices and technology systems that could impact data sharing and data governance.
- Criminal justice records across criminal justice agencies are stored as different units of analysis, which makes centralizing records more complex, especially when merging, linking, and restructuring files from the Department of Public Safety (PSD) (which looks at individuals who enter their facilities) and the Hawai`i Criminal Justice Data Center, Department of the Attorney General (HCJDC) (which houses arrest records for individuals) to merge with the Judiciary (which houses information on court cases that are organized by case numbers and not individuals).
- Some data can be pulled from agency databases as structured data fields, however many are text based and even more challenging are the unstructured text fields that include long comments or court minutes.
- Extensive manual labor from several staff across pretrial agencies and CJRI are required to create a centralized source of pretrial data for research since staff must extract multiple tables from each IT database, share files across agencies, and then merge, link, and transform several fields of information into data that can be used for required pretrial statistics.

As the CJRI director interviewed several states and jurisdictions embarking on similar database projects, it became clear there was no one solution to create the pretrial database and reporting system. There are as many solutions in existence as jurisdictions embarking on these projects. While many barriers are similar across states, solutions depend on the rules, laws, technology, resources, policies, and organizational culture of these organizations. After surveying different solutions, CJRI categorized three different approaches to this work and made a recommendation to the CJRI board. The board

reviewed the costs and benefits of the different approaches, and agreed with the recommendation from the CJRI director to pursue resources for a technological solution that would extract data from all three agencies and centralize it into a data warehouse. A feasibility study was conducted in the Fall of 2022 to verify the technical requirements and estimate a budget and timeline for creating a pretrial database and reporting system. This informed the appropriation request in Act 147.

In summary, CJRI reviewed data sources across the state to develop a technical plan to create the pretrial database and reporting system. This was done in collaboration with criminal justice stakeholders which house the three main sources of pretrial data - PSD, HCJDC, and the Judiciary. The aim was to identify a solution that would address the barriers to reporting out on the pretrial system in a timely and efficient way. Act 147 provides funds and resources for CJRI to carryout this solution.

Due to the complexity of IT systems, databases, software, and criminal justice decisions involved in creating a statewide pretrial database and reporting system, the next page provides a conceptual overview of the project.

WHAT IS HAWAI`I'S PRETRIAL DATABASE AND REPORTING SYSTEM?

Explaining the creation of the "centralized statewide criminal pretrial justice data reporting and collection system" (HRS § 614-3)

Agencies use case and records management information systems in the pretrial system

Criminal justice agencies use several information systems to collect and store information on individuals entering the criminal justice system. These systems store information on court cases, arrest records, and more.

Three agencies store most statewide information on the pretrial system

The CJIS system (Department of the Attorney General) contains arrest records, JIMS (Judiciary) includes information on court cases and court decisions, and OffenderTrack (Department of Public Safety) includes information on people entering and exiting jails.

Records on cases and people must be linked across the three systems

Each agency uses a unique identifier on cases or people to store information in their systems, and these identifiers must be used to link court records, arrest records, and jail records together in order to analyze system trends or evaluate impacts.

IT tools will create data pipelines for each agency to submit data into a centralized data warehouse

The pipelines will link records and store them in one centralized data warehouse, making it possible to create one database of pretrial information without requiring additional data entry.

Centralizing data will provide more effective and efficient data capacity for reporting

By linking records in one location, a platform will be created to build dashboards for regular reporting on metrics and will establish a central source of datasets to extract for evaluation and analysis.

TABLE OF CONTENTS

PROVIDING PROGRESS UPDATES ON ACT 147

This report is organized with both a high level list of activities and a summary to provide updates to the Legislature on the development of the pretrial database and reporting system. Several activities related to the project are organized into thematic sections. Much of the work is sequential, therefore as tasks are reviewed further down the list, fewer have been initiated. For example, reporting cannot occur until the centralized data warehouse is created and the data pipelines are built. The list of activities are based on project deliverables from contracts, but written at an overview level to capture the main components of the system. If there is no progress update such as "in progress" or "ongoing," then the task has not been initiated. A written summary is provided to capture context or to explain future activities for the project.

This project cannot be accomplished without the full cooperation from each of the agencies CJRI is collecting data from for the pretrial database and reporting system. This project was planned in consultation with operations, IT, research, and other staff across the three agencies to ensure it was a feasible approach. CJRI is grateful for their continued support of this project, including their testimony in support of House Bill 68 during the legislative session, which became Act 147. CJRI staff would like to acknowledge the continued involvement of PSD, HCJDC, and the Judiciary to make this project a priority. This includes meeting with IT vendors, sharing datasets, answering questions about pretrial operations, and other activities requested of them. CJRI is grateful for this continued support.

PROGRESS UPDATE

Administrative and Initial Development Activities	In Progress	Ongoing	Completed
Executed contract			●
Met with IT partners to establish project plans and timelines			●
Met with three primary agencies to introduce data transfer protocols and expectations			●
Developed draft data governance agreement			●
Meet with criminal justice peers in other states with similar centralized data and reporting systems		●	
Create advisory groups of subject matter experts for ad hoc consultation	●		

Summary of Progress

A contract was executed in September 2023 for the pretrial database and reporting system through the state procurement office's price list and includes four IT partners. The contract is for two years, which is the estimated time it will take to complete the work. After the execution of the contract, CJRI staff met with all contracted parties to initiate the work. This included bringing the IT partners together with the three criminal justice agencies participating in the project to coordinate work and data sharing across the Judiciary, PSD, and HCJDC.

The project is comprised of several deliverables from different IT partners, which together, will create several components of the pretrial database and reporting system. These components include 1) a data ingestion process that merges, links, and stores records into 2) a centralized data warehouse while creating 3) a streamlined process for data transformation to turn information into variables for statistical analysis, in addition to establishing 4) a platform to host dashboards to report out on key metrics frequently and 5) a site to store data into datasets for more complex evaluation research.

Data governance and data sharing policies are critical to projects like the pretrial database and reporting system. Each agency has different rules around data sharing and varying levels of confidential information that must be handled through secure protocols. While there are current data sharing rules in place as the system is developed, a draft data governance policy was created to initiate conversations for the final policy. With the Extract, Transform, Load (ETL) tool that CJRI has selected, there are various ways to set-up data sharing processes and to apply different security rules. The data governance policy will be reviewed and revised with feedback from all three agencies as the tool is developed to meet each agency's needs. For example, some data may need to be masked before it is ingested into the centralized warehouse, while other data may need to have certain identifiers removed from the datasets before they are shared. Data governance will be developed based on the individual needs of each agency's data sharing rules, and the uniqueness of the data.

Additionally, CJRI is engaging subject matter experts throughout this process. This includes connecting with other jurisdictions who have engaged in similar projects. For example, CJRI and their IT partners met with a site in New York City to learn more about their process for coordinating IT and research projects strategically. CJRI is also developing subject matter expert groups to consult on issues like data governance, research, operations, and other areas that can benefit from consultation as the system is developed.

| Identification of Data and Data Transfer Process	In Progress	Ongoing	Completed
Obtain sample dataset extracts from all three agencies		●	
Create data map demonstrating linking of data records		●	
Develop calculations for key performance indicators		●	
Create data codebook		●	

Summary of Progress

Before any datasets can be analyzed or metrics can be reported, there must be a process established to ingest the data into a centralized data warehouse, which requires a significant amount of background work. Several datasets from all agencies must be shared in order to create data maps that demonstrate what identifiers across datasets will link records effectively, as well as what data elements must be used to transform and calculate information into variables and metrics. Agencies do not have data codebooks, therefore CJRI is creating new data background materials that will benefit all researchers in the state. In many cases, several files from the three agencies are needed to create a complete picture on someone's trajectory throughout the pretrial period. This mapping of records and fields is necessary to ensure data pipelines are created that reflect these maps.

All three agencies have been very engaged and supportive of this project. Staff at PSD, HCJDC, and the Judiciary have shared data when asked, taken the time to answer questions about data as well as provided insight into the operations the data is capturing. The staff at CJRI are grateful for the cooperation provided by staff members across these three agencies.

Why is data mapping important for a centralized data warehouse? As someone progresses through the criminal justice system, decisions points filter people and cases out like a funnel. Figure 1 below demonstrates this.

Figure 1: Criminal Justice Funnel for Pretrial Data
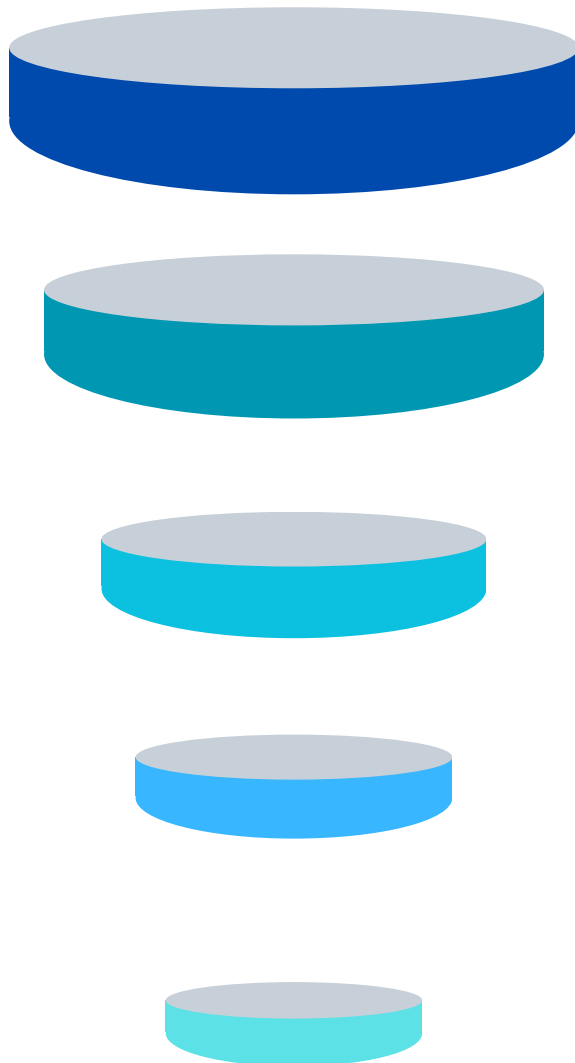
Arrests
Arrests occur with local police, and arrest data are aggregated in CJIS.

Charges
Prosecutors must file charges to initiate a case, which is tracked in Judiciary JIMS data. Not all arrests are charged.

Jail bookings
Not all individuals charged will be brought into a PSD facility, whether for booking or pretrial detainment.

Adjudications
Pretrial ends when a case is adjudicated, either dismissed or sentenced.

Convictions
Not all adjudications are convictions, an individual must plea or be convicted by jury. Some individuals may have their cases dismissed or be found innocent.

As records are matched across data from HCJDC, PSD, and the Judiciary, identifiers are used to link these records. In some cases it is a unique identifier for a person, other times it is a unique identifier for an arrest report or a court case. Not all records will have matches as cases filter out of the system. And for some metrics, it requires all three data sources to calculate metrics for pretrial.

Creation of the Centralized Data Warehouse	In Progress	Ongoing	Completed
Hold meetings with each agency's IT department to understand technological capabilities		●	
Install ETL software	●		
Establish SFTP connections for each agency to share data	●		
Develop data pipeline strategy	●		
Receive training on ETL			
Receive training on data warehouse			
Finalize ETL architecture to promote most efficient method of data sharing long-term			
Finalize security policies and procedures			

Summary of Progress

The ETL provides different ways for each agency to share data with the data warehouse. For now, an SFTP connection will be established for each agency instead of sharing through encrypted emails or flash drives. The final data sharing process will be finalized with feedback from each agency based on their technology and data security needs. For instance, the ETL can be set-up with a direct connection that would create an automatic process for data sharing with CJRI. Once finalized, staff will receive training on the use of the tool and the data warehouse. Data governance policies will be updated. The database will be setup to make data sharing efficient, while respecting the rules governing the sensitive data.

Development of Reporting Process	In Progress	Ongoing	Completed
Author data visualization guide for dashboard and reporting methodology and design	●		
Receive training on software for reporting and dashboards			
Develop draft dashboards with software for main stakeholder groups			
Provide demonstration of priority metrics for feedback			
Conduct stakeholder engagement to receive feedback on reporting	●		

Summary of Progress

Most of the work related to the development of the reporting process will be initiated later on in the project. The data must be ingested into the centralized data warehouse first before dashboards can be created. The intent behind HRS § 614 is to provide metrics and research that will help inform policy recommendations. As a result, CJRI recommended that the pretrial database and reporting system help improve the efficiency of data collection in order to provide more timely reporting. The ETL and data warehouse are critical for achieving this goal. With an efficient process to ingest data into the data warehouse, dashboards can be updated at least monthly on pretrial trends in addition to providing access to datasets that allow CJRI research staff to analyze data more quickly. The bulk of time that research staff spend on data projects is due to data cleaning, which includes the linking and merging of records, and the transformation of information into variables for quantitative analysis. By leveraging technology to prepare datasets, research staff will be able to analyze data

more quickly to inform policy decisions.

In the meantime, CJRI staff are creating a data visualization guide to standardize reporting and outline a research methodology for pretrial metrics. It is important that the appropriate data visualizations are used for reporting on pretrial metrics, such as selecting between histograms or pie charts to convey findings about the pretrial system. Additionally, many publicly available dashboards lack information that allow consumers to understand the main takeaways presented in the dashboard. Information must also be presented about the methodology behind metrics including the source of data, the sampling approach, the calculations, and other factors that can impact the limitations of interpreting the metrics. CJRI staff are developing a guide to address this with the assistance of an intern majoring in both criminal justice and data science, analytics, and visualization provided by Chaminade University of Honolulu.

In addition to the data visualization guide, CJRI staff will share draft dashboards and reporting templates with key stakeholders to ensure the information is clear and digestible. This encompasses a wide range of stakeholders including legislators, criminal justice agency staff, and the public. CJRI will provide updates on their website about opportunities to provide feedback on the new system as they become available.

Finalizing the Pretrial Database and Reporting System	In Progress	Ongoing	Completed
Execute final data governance agreement			
Finalize data sharing MOUs for new system			
Revise dashboards based on feedback from stakeholder engagement			

Finalizing the Pretrial Database and Reporting System (continued)   In Progress   Ongoing   Completed

Make dashboards public on CJRI website

Create new section of annual report
summarizing pretrial metrics and reporting

Summary of Progress

The last remaining list of tasks are related to finalizing the policies and technical features for the pretrial database and reporting system. CJRI is working with staff across different departments in all three agencies including IT, operations, staff attorneys, and others. The database must be created in a way that follows different agency needs. For example, each agency has different IT capabilities and different rules around data sharing. Additionally, some policies will be adjusted to reflect the technical aspects of the new system such as revising data governance policies to reflect the final database architecture established for data sharing with each agency. Finally, the database and the reporting process will incorporate feedback from different stakeholders to create a system that translates pretrial metrics and research in a digestible way to inform statewide pretrial policy.

Additional Considerations

Staff support: As part of Act 147 (2023), funds were allocated to create the new system, while also providing support to make the project successful. The law created two temporary IT positions - one for PSD and one for the Judiciary. Both have been established and are in the recruitment phase. These positions will provide their agencies with the capacity to front-load a significant amount of work for this project. The new law also provided CJRI with a new permanent position to support the database long-term. This position is being created with feedback from technical experts in order to recruit an individual for the position that would limit the need for external assistance in the future.

Database security: A substantial amount of pretrial-related data can be found in public records, such as arrest reports or court records. However, many of these systems also include confidential information associated with these records. As often as possible, CJRI is extracting data that is public, while omitting and/or not utilizing confidential information, as well as personally identifiable information (PII) such as social security numbers and home addresses. The ETL and cloud storage will have policies and procedures on data security, and will be set-up to limit the storage of PII. For example, if it is needed for matching records, the ETL can be set-up to match records with PII but remove the PII before it is stored in the cloud-based database. Additionally, the data warehouse that will be utilized is being used in many states and has security protocols appropriate for confidential data, including the CJIS data housed by HCJDC.

Even if records are public, social science researchers strive to keep people anonymous and to protect them from additional scrutiny. CJRI is working with their IT partners to make it difficult for a specific individual to be identified even if the information is public. The system is designed for reporting at the aggregate and is not a system to query individuals to piece together their entire criminal justice record. CJRI data sharing MOUs address these concerns, and are modeled after research data sharing agreements that cover sensitive information.

Long-term planning: As CJRI staff have assisted with other research requests, they encounter the same barriers to criminal justice research regardless of the topic. Criminal justice data is disconnected, and many operational decisions are made in silos. Because of this, it is difficult for agencies to turn around information on other policy questions in a timely manner. If the Legislature wanted to learn more about probation or parole, similar barriers exist as those in the current pretrial data landscape. The new pretrial database can serve as a model to help the state strategize about solutions to collect criminal justice data beyond the pretrial system.

Finally, it is not enough to create a pretrial database and reporting system

to change policy. The Legislature, agencies, and the public must embrace data driven decision-making. This is why CJRI staff have invested time with agencies to build trust in the data collection and reporting process among stakeholders. This includes working with agencies to ensure CJRI research staff know the strengths and limitations of the data, and developing a reporting process that is objective with documented methods that are transparent and clear. CJRI staff is coordinating with other statewide efforts to modernize data and build data capacity for the state. For example, this includes meeting with the new Chief Data Officer, Office of Enterprise Technology Services, or collaborating with local organizations and universities to bring more expertise to the project. These partnerships can help facilitate discussions with stakeholders on improving data driven decision-making or bring together other peers leading this work in other state agencies.

OUR ORGANIZATION

CJRI STAFF

Erin E. Harbinson, PhD
Director

Aerielle Reynolds, MSCJA   **Samuel Choi, PhD**
Research Analyst   Research Analyst

Mariah A. McCaskill
Secretary

Pamela Oda
Undergraduate Research Intern

I ka nānā no a ʻike.

By observing, one learns.
-ʻŌlelo noʻeau

Through observing, or researching, Hawaiʻi's criminal justice system, CJRI is dedicated to helping stakeholders, lawmakers, and the public learn more about Hawaiʻi's criminal justice system.

BOARD MEMBERS

Judge Matthew J. Viola
CJRI Board Chair
Senior Family Court Judge
Judiciary

Nicole C. Fernandez
Corrections Program Specialist
Department of Public Safety
Governor's Office

Rep. Scot Z. Matayoshi
District 49
Hawai`i House of Representatives

Peter Wolff
Federal Public Defender (Retired)
Hawai`i Senate

Francis Young
Intake Service Centers Division
Administrator
Department of Public Safety
(Term starting August 1, 2023)

Shelley Harrington
Department Human Resources Officer
Department of Public Safety
(Term ending July 31, 2023)

The staff at CJRI could not accomplish their work successfully without the expertise of the board. Each of the board members brings valuable knowledge from their respective roles and experience across the criminal justice system and the policymaking realm. The criminal justice system is wide-ranging, and the board is essential in helping prioritize projects and providing feedback on ways to communicate research. Their collective experience has improved the work of CJRI in several ways. The CJRI staff thank the board members for their ongoing work and support.

ENDNOTES

1. *Hawaiʻi Criminal Pretrial Reform: Recommendations of the Criminal Pretrial Task Force to the Thirtieth Legislature of the State of Hawaiʻi* (2018): https://www.courts.state.hi.us/wp-content/uploads/2018/12/POST_12-14-18_HCR134TF_REPORT.pdf
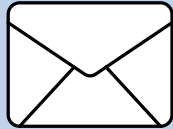
ACKNOWLEDGEMENTS

CONTACT INFORMATION

Criminal Justice Research Institute
The Judiciary - State of Hawai`i
417 South King Street
Honolulu, HI
96813-2943

808 - 539 - 4881

CJRI@courts.hawaii.gov

https://cjrihawaii.com/

https://www.courts.state.hi.us/criminal-
justice-research-institute-cjri