

# Modernizing California Pretrial Data Pipelines for Responsible Risk Assessment

Allison Day  
Anil Dewan  
Karissa McKelvey



**ASPEN TECH POLICY HUB**



*Photo by Siyan Ren on Unsplash*

## Executive Summary

California counties increasingly rely on algorithmic risk assessment tools to help judges make decisions about whether to detain or release pretrial detainees. At stake is the liberty of individuals not yet convicted of crimes, and the reduction of the massive pretrial jail population—a primary driver of mass incarceration.<sup>1</sup>

For California county pretrial agencies to use risk assessment tools responsibly, they must have consistent data collection practices, modern data management systems, technical expertise, and data access, as well as clearly defined metrics for success. Currently these capabilities are outdated and fragmented, and technical development is under-resourced or non-existent. California's risk assessment systems urgently require modernization to improve overall pretrial outcomes.

Counties using algorithm-based methods should follow best practices to improve their pretrial risk assessments, including:

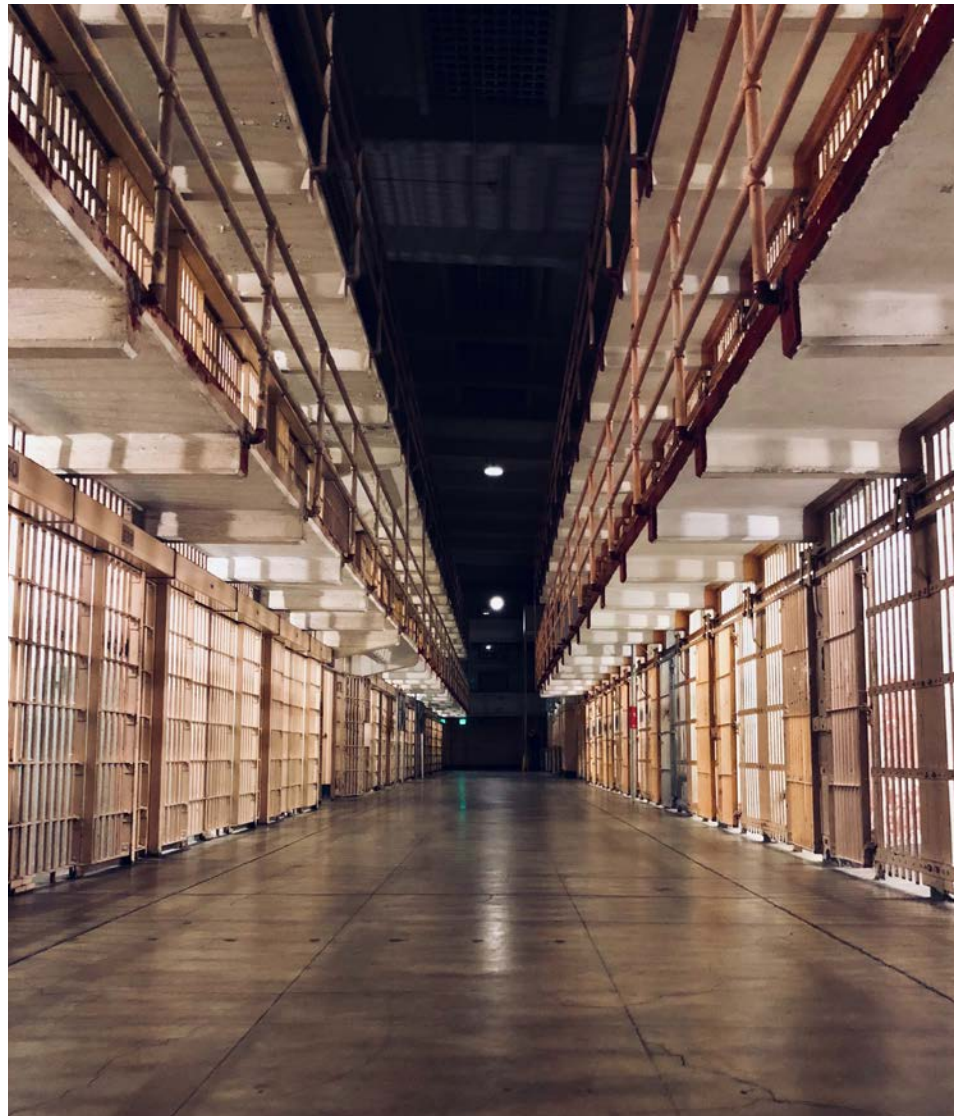
- ▶ Pulling criminal background data from multiple databases efficiently to perform risk assessments quickly;
- ▶ Validating the quality of pretrial risk assessment predictions frequently and ensuring they achieve desired local outcomes;
- ▶ Extracting, cleaning, normalizing, and aggregating data in a consistent manner at the county and state levels;
- ▶ Analyzing outcomes and sharing data visualizations with justice partners at the local, county, state levels, as well as with the Judicial Council of California; and
- ▶ Publishing outcomes and validation data publicly to ensure transparency.

Without the technology and staff necessary to carry out these practices, reduction in pretrial jail populations will be stalled and pretrial agencies will continue to conduct pretrial risk assessments without critical data-driven insights about the effectiveness and fairness of the process.

This report makes four key recommendations for the Judicial Council of California to implement:

- ▶ Modernize the end-to-end pretrial data pipeline (extraction, cleaning, normalization, aggregation, analysis, and visualization) using customizable applications that easily scale for both county- and state-level data pipeline needs;
- ▶ Hire software engineers and data analysts capable of building secure, flexible software tools using agile, iterative practices for increased security, responsiveness to bugs, and implementation of new features;
- ▶ Partner with modern software companies to vastly reduce the time and cost required to implement solutions; and
- ▶ Default to open-source software to reduce costs and increase security, transparency, and ownership of data.

*Photo by Carles Rabada on Unsplash*



Through the use of modern data applications, the pretrial risk assessment process will be more accurate, speedy, transparent, and just. Successful modernization will lead to a variety of benefits, including a reduction in pretrial jail populations, faster release of low-risk individuals, and safer communities.

## Background of Risk Assessment in California

### Why Pretrial Reform?

Reducing excessively large pretrial jail populations is essential to ending mass incarceration and criminal justice reform. The Bureau of Justice Statistics estimates that 63 percent of the 720,200 people held in U.S. jails in 2015 were held pretrial, before they had been convicted of a crime.<sup>2</sup> Effective pretrial risk assessment programs can reduce jail populations, reduce the length of incarceration, and minimize life disruption caused by incarceration. In particular, contemporary statistical pretrial risk assessment tools can speed the pretrial process and help inform judges' decisions to release or detain.

### Why Focus on Data?

Pretrial risk assessment services have existed in California since the 1960s. Technology systems and data collection practices have changed significantly over the last 50+ years and pretrial risk assessment tools have evolved from simplistic tools to the sophisticated algorithmic, data-driven products used today. Despite their long history, it is still extremely challenging to gauge the impact of pretrial risk assessment tools, both locally and statewide.<sup>3</sup> Challenges exist on numerous levels, including lack of algorithmic transparency, inconsistent implementation, and infrequent tool validation.

Numerous experts, including criminal justice reform advocates and AI and statistics researchers, have called into question the effectiveness of pretrial risk assessment tools in reducing incarceration rates, one of the fundamental goals of pretrial reform. The reports also raise issues of bias in the models and the data used to build them, making it even more critical to evaluate their effectiveness in real-world scenarios.<sup>4</sup> Advocates have repeatedly called for greater transparency



Photo by Bill Oxford on Unsplash

into the data sets used to train the tools, and have asked the courts to frequently and consistently publish outcomes and data on potential bias.<sup>5</sup> Furthermore, to ensure the utility of a given pretrial risk assessment tool in a given jurisdiction, experts recommend tailoring risk estimates and pretrial decision-making policies to jurisdiction-specific failure rates over relatively recent timeframes.<sup>6</sup>

This growing body of research has incentivized the California State Legislature to propose SB-36, which imposes data collection and reporting requirements on the Judicial Council regarding the usage of risk assessment tools across the state. Pending the outcome of SB-36, counties are already heeding recommendations about the periodic evaluation and auditing of pretrial risk assessment. Although some counties have improved or are beginning to improve their processes and pretrial information systems, their progress continues to be significantly limited by resourcing and lack of appropriate technical expertise.

This report articulates the challenges with current data practices, and with management and oversight of pretrial risk assessment systems. As an alternative to traditional enterprise procurement, we offer recommendations for quickly and cost-effectively building customizable open-source software (OSS) that can integrate with existing legacy systems across the state to produce the analytics, dashboards, and reporting necessary for data-driven pretrial reform.

## Challenges for Data Collection and Reporting

### Overview of Challenges

Current data collection and reporting systems are broken at the county level. Consequently, effective state-level analysis and oversight is impossible. In this section, we offer a breakdown of current challenges and a review of systems used by three California counties. In the next section, we offer specific recommendations to the Judicial Council to address the issues described.

Our research included extensive interviews with AI and pretrial reform policy experts about pretrial assessment systems and data from published papers. We conducted multiple interviews and held in-depth discussions with four counties across the state about their current data collection and sharing practices and the challenges they face. We analyzed the work and products of four organizations and interviewed leaders and engineers engaged in developing criminal justice software solutions, for California and nationally. We also had conversations with the Judicial Council about their plans for the \$75 million appropriation from the Budget Act of 2019. Finally, we used the data collected and our backgrounds as technologists in industry and the public sector to analyze the challenges and develop the recommendations presented below.

### Findings

Below is an overview of specific data management challenges encountered by pretrial agencies, based on our interviews with pretrial stakeholders from several California counties.

#### Manual data collection from antiquated systems is inefficient

- ▶ County pretrial programs must be able to view, extract, and merge data from between 4-5 judicial partners, including:
  - ▶ Sheriff's jail management system;
  - ▶ Court management system;
  - ▶ National Crime Information Center Database; and the
  - ▶ Local pretrial services database.
- ▶ Access and permissions for each system are managed separately.

- ▶ Data is stored in outdated legacy systems—some over 20 years old—that require manual extraction of data points onto paper or spreadsheets, and that can only be collected by calling agencies by telephone.
- ▶ The time-consuming manual process required to gather data for assessment burdens staff and delays processing, keeping people in jail longer.
- ▶ Legacy systems lack critical data export, integration, aggregation, analysis, and visualization capabilities.

**Technical strategy is not centralized**

- ▶ The burden is on individual counties to develop and implement methods for data collection, analysis, and reporting results.
- ▶ The budget for technical solutions and staffing competes with other priorities.
- ▶ There is a missed opportunity for sharing technical resources and tools between counties to support common needs.

**Lack of data expertise forestalls system improvements**

- ▶ Many counties do not have the resources to employ or contract technical experts to build and maintain their data systems.
- ▶ Fears that the transmission of confidential data is vulnerable to hacking—even when it is legitimate and can be achieved securely—can delay or block potential improvements in data practices.
- ▶ Leadership and staff lack education on contemporary data security and privacy practices, such as encryption standards, secure data transfer and storage, and secure data access, that can make sharing safer and easier.
- ▶ Stakeholders block or limit the release of pretrial outcomes data due to fear of being held accountable for mistakes in risk assessment and release decisions, despite calls by experts and advocates for greater transparency.

**Slow and costly enterprise procurement is viewed as the only solution**

- ▶ Existing data collection practices require expensive multiyear development and maintenance contracts.
- ▶ The use of proprietary software offers limited accountability and limited control of new feature requests, bug fixes, or data management.
- ▶ The IT infrastructure costs of maintaining standalone data centers are extremely high compared to commercial cloud services.
- ▶ Cost can be prohibitive for small or under-resourced counties.
- ▶ Counties continue to use outdated infrastructure and manual data pipelines that are prone to failure.
- ▶ Front-loaded software contracts are antagonistic to the agile, iterative practices used by modern software companies.
- ▶ Fixes and feature requests are subject to slow response times.



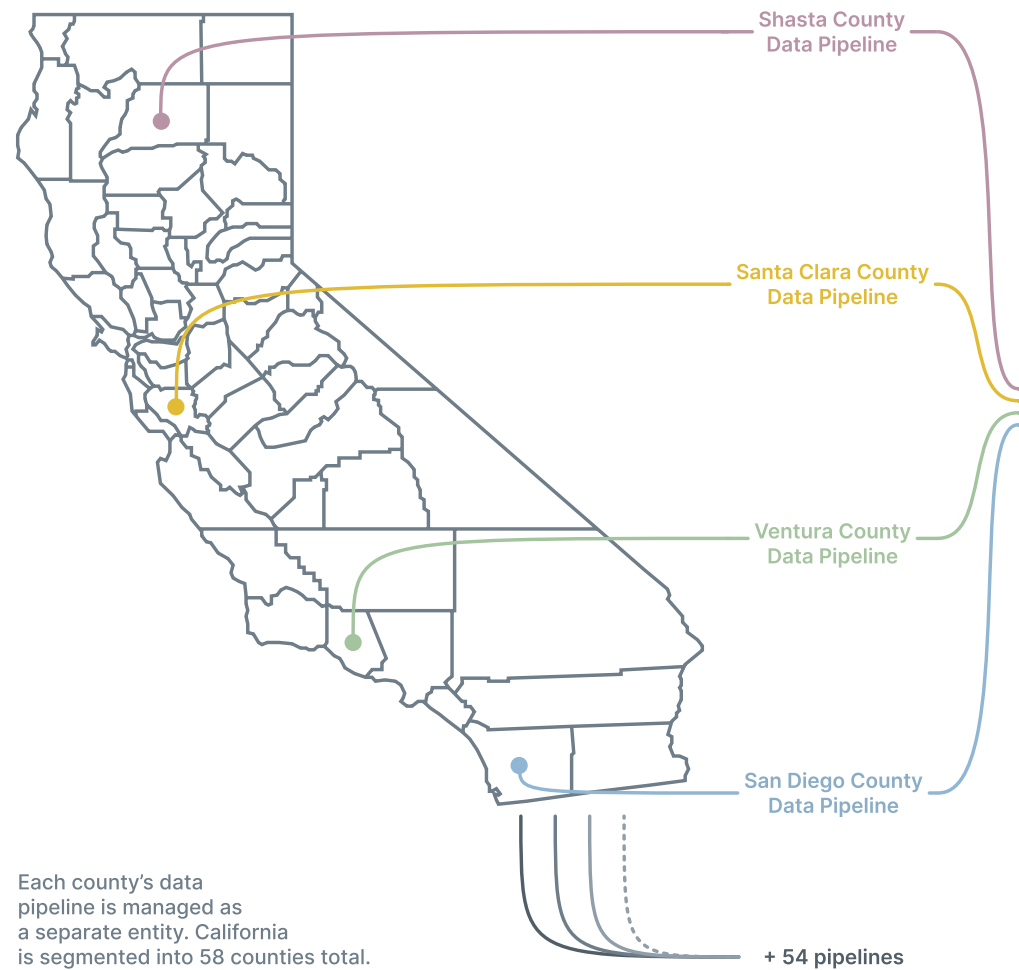
Photo by Markus Spiske on Unsplash

The “Current System by County” diagram presented below depicts how a typical data process of extraction, collection, and analysis works. The current approach treats each county as a separate entity responsible for implementing its own technical strategy, and leads to duplication of counties’ efforts to build software tools and data processes. This results in a highly fragmented and non-scalable approach to data collection, sharing, and validation, and requires hiring technical resources at each county, which is cost-prohibitive for all but the largest counties in the state.

A typical government-driven approach to renovating IT systems is to procure proprietary all-in-one solutions with limited licensing rights and expensive support agreements. If the Judicial Council or individual counties pursue this route, the result will be expensive, multiyear efforts to replace legacy systems. Operating and maintaining these systems would require additional support agreements or hiring technical personnel in each jurisdiction.

Because counties have limited budgets and are under critical pressure to fix pre-trial data systems quickly, we recommend investing in modern data tools built by software companies that operate at a fraction of the cost and time frame. The following section reviews two California counties that used this method successfully, as well as a county that continues to operate its pretrial services using spreadsheets and other manual processes. These case studies are based on interviews with people in charge of operations and technology at pretrial agencies, as well as personnel involved in the day-to-day execution of pretrial services in these counties.

## Current California Pretrial Data Pipeline



### OVERVIEW OF CURRENT CALIFORNIA PRETRIAL DATA ECOSYSTEM

Current technical strategy is based on geography

Each county exists as a separate entity, responsible for developing and implementing its own software systems and data pipelines

Counties typically need to view, extract, and merge data from 4-5 judicial partners. This effort is currently repeated manually, or with some automation, in 42 out of 58 counties across the state

Local counties and the Judicial Council hire their own staff, resulting in a duplication of skills and personnel across the system

Most counties procure expensive and proprietary software solutions, limiting their ability to share code and solutions across boundaries

#### ADVANTAGES

Each jurisdiction creates its own highly customized system

#### DISADVANTAGES

Extremely high cost of implementation and management

Extremely high cost of infrastructure and maintenance of data centers

Extended timelines to implement data pipelines as each county has to work on its own

Human resource intensive

Fails to leverage the benefits of technology at scale or share solutions with each other due to proprietary software

Creates a highly fragmented system that is prone to failure

New features and upgrades cannot be launched system-wide

Shasta Process

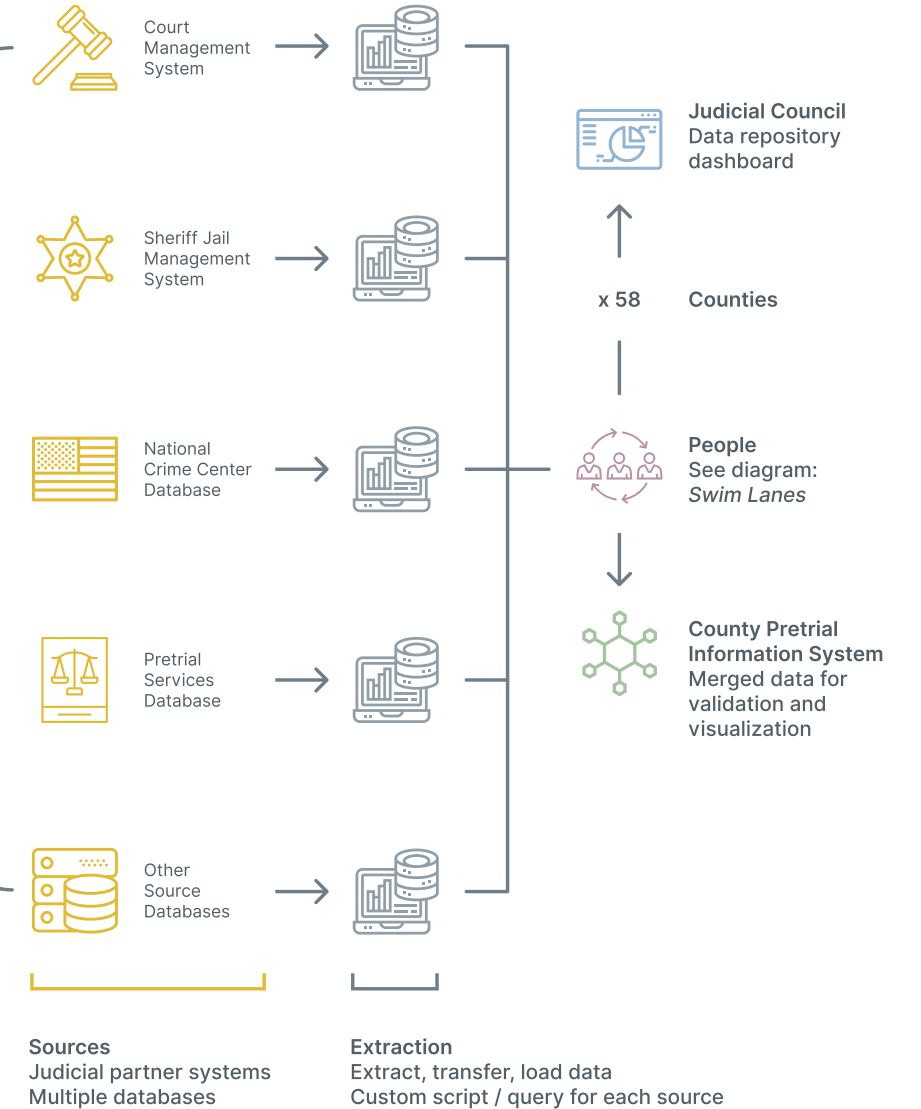
Santa Clara Process

Ventura Process

San Diego Process

+ 54 processes

### TYPICAL COUNTY-LEVEL DATA PROCESSING: Source, Extraction, Destination



## County Case Studies

San Francisco and Santa Clara are two examples of counties with robust pretrial programs that have invested in sophisticated pretrial information systems that can clean data, test validity, analyze results, and present metrics. Del Norte County, by contrast, exemplifies the case of a smaller county with extremely limited resources where technical improvements have taken a back seat to the immediate needs of client case management, court advocacy, and program coordination.

San Francisco is a relatively resource-rich county, in terms of both funding and proximity to Silicon Valley. San Francisco partnered with the California Policy Lab (CPL) to build its pretrial data management systems. CPL built tools to pull and aggregate data from across several agencies to feed into pretrial risk assessment. It also created a data visualization dashboard that allowed the San Francisco Courts, the Sheriff's office, the county's nonprofit pretrial agency, and others to easily analyze trends and validate their risk assessment tool. CPL also defined metrics to measure outcomes and worked with San Francisco to perform analysis and report on pretrial trends, using the improved data visibility enabled by CPL's tools.<sup>7</sup>

Similarly, Santa Clara County built its own pretrial data system, the Pretrial Online Production System (POPS),<sup>8</sup> to fulfill similar functions. Santa Clara has built customized, advanced, and automated pretrial information systems that pull data from partners and generate metrics on a daily basis. A relatively well-resourced county, Santa Clara used in-house staff to build its data system.

In contrast, a much smaller county like Del Norte has a single staff person who serves as the "one-stop shop" for pretrial services. Del Norte County could not afford to purchase its own iteration of the Probation Department's hefty and rigid enterprise system for its case management and data needs. The county instead returned to using time-consuming manual methods and spreadsheets following a brief pilot with the enterprise system. Del Norte reverted to the manual system in part because needed fixes and features were never addressed by the software company in charge of the system during the pilot period.

Del Norte is not unique in size and resourcing. In 2015, Californians for Safety and Justice reported that 20 out of 58 California counties had fewer than three full-time employees (FTE) in their pretrial services division, and a majority (35) had less than 10 FTE.<sup>9</sup> As a result, many of California's counties do not have sufficient resources to build their own pretrial data and reporting systems. Smaller counties simply cannot afford staff for the technical functions of data analysis, operations, and engineering, nor do they have the budget to purchase traditional enterprise software.

## Case Studies: California Counties

San Francisco	Santa Clara	Del Norte
Population: 884,363 (2017)	Population: 1,938,000 (2017)	Population: 27,470 (2017)
Urban?: Yes	Urban?: Yes	Urban?: No
Working Group?: Yes	Working Group?: Yes	Working Group?: Yes
External Data Partner?: Yes	External Data Partner?: No	External Data Partner?: No
Probation Runs Pretrial?: No	Probation Runs Pretrial?: No	Probation Runs Pretrial?: Yes

### Key Features

<p><b>1   Quarterly Pretrial Stakeholder Working Group</b></p> <ul style="list-style-type: none"> <li>Build mutual understanding and consistency on risk assessment</li> <li>Regular review of judge concurrence rates</li> <li>Resources to validate PSA</li> <li>Update Decision Making Framework</li> <li>Train new staff</li> </ul>	<p><b>1   Investment in Risk Assessment Validation</b></p> <ul style="list-style-type: none"> <li>Shifting from VPRAI to PSA after careful analysis</li> <li>Well-defined metrics and regular review</li> <li>High judge concurrence rate with risk assessment recommendations</li> </ul>	<p><b>1   Small and Scrappy Organization</b></p> <ul style="list-style-type: none"> <li>One pretrial coordinator handles everything, including pretrial case management, risk assessment, data management and analysis, court advocacy, and reminders</li> <li>Most processes are done manually or with spreadsheets</li> <li>Time consuming to run rap sheets, interview clients, manage data and records, etc.</li> </ul>
<p><b>2   Collaboration with the California Policy Lab</b></p> <ul style="list-style-type: none"> <li>CPL: Independent academic agency</li> <li>De-silos data across pretrial agencies</li> <li>Builds dashboards to allow easy visualization, analysis and reporting</li> <li>Helps SF be ready for SB-36</li> </ul>	<p><b>2   Bail Release Group</b></p> <ul style="list-style-type: none"> <li>Dedicated to pretrial</li> <li>Includes community partners in domestic violence, immigrant rights, racial equity &amp; tech advocacy</li> </ul>	<p><b>2   Lack of Funds for Case Management System</b></p> <ul style="list-style-type: none"> <li>Trial run with Probation Dept's case management and data system</li> <li>Vendor wanted to sell the entire system when only certain functions were needed</li> <li>Requested fixes were never made</li> <li>Assessment process way too slow - can't keep up with jail population</li> </ul>
<p><b>3   Nonprofit-run Pretrial Services</b></p> <ul style="list-style-type: none"> <li>Wraparound case management</li> <li>Holistic needs assessments</li> <li>Peer-based services: Many staff have been through the system</li> <li>Nimble enough to adapt to changing populations</li> <li>Special grants (ie HIV, homeless populations) to increase effectiveness</li> </ul>	<p><b>3   POPS</b></p> <ul style="list-style-type: none"> <li>In-house pretrial data management system</li> <li>Built and maintained by county IT department</li> <li>Dashboards with key daily performance metrics</li> <li>Ability to validate risk assessment and analyze overall outcomes</li> <li>Built with non-open source software</li> </ul>	<p><b>3   Underresourced Pretrial and Jail Overcrowding</b></p> <ul style="list-style-type: none"> <li>Small counties cannot afford costly case management systems and technical staffing</li> <li>Pretrial risk assessment is cut out of the equation when it doesn't have enough resources to operate</li> <li>Del Norte jail forced to make release determinations without risk assessment</li> </ul>

## Recommendations and Resources

### Why now? The Judicial Council of California and pretrial programs have funds to expand services

The Budget Act of 2019 appropriated \$75 million in one-time funding to the Judicial Council of California for distribution to counties to fund the implementation, operation, and evaluation of programs or efforts related to pretrial decision-making in at least 10 courts. On August 9, 2019 the Judicial Council awarded grants to 16 California counties. Per a release by the Judicial Council “the projects **aim to increase the safe and efficient release of arrestees before trial; use the least restrictive monitoring practices possible while protecting public safety and ensuring court appearances; validate and expand the use of risk assessment tools; and assess any bias.**”<sup>10</sup> The majority of counties funded by this grant plan to “expand or greatly enhance” pretrial risk assessment as a key strategy to reduce jail populations and reform pretrial services. For example, Los Angeles County is developing an innovative two-step pretrial assessment process, Sonoma County has set a target of completing pre-arraignment releases within 12 hours of booking, and Calaveras County has implemented a new pretrial program with the county’s probation department.

The Judicial Council held back approximately 10 percent of the appropriation to support technical integration and development with each county but did not specify how it plans to use those funds.<sup>11</sup> With the increased emphasis and expansion of risk assessment, effective data practices are crucial, which the Council failed to prioritize in its awards. We therefore recommend that the Council use these funds to:

- ▶ Develop new tools to hasten pretrial reform and maximize the ability for all California counties to responsibly use algorithmic risk assessment tools, and
- ▶ Deploy modern data collection, analysis, and sharing practices so the program can be governed more effectively.

#### We offer four specific recommendations based on our research and expertise:

- ▶ **Build modular data applications:** Address the need for data pipeline improvements across California counties by building software that can be customized to integrate with diverse data systems.
- ▶ **Hire experienced software engineers and data analysts:** Hire technical personnel whose product development skills and technical expertise will ensure they are building streamlined, secure, and flexible software using non-proprietary tools.

- ▶ **Work with modern software development companies:** Contract with companies that use an agile, iterative approach to development and open-source software as a matter of practice.
- ▶ **Use open-source software (OSS):** Ensure security, transparency, and affordability of pretrial data systems through OSS and implement a policy for its use.

### Build modular data applications that can be shared across multiple counties

Data applications consist of software programs that can extract, clean, normalize, aggregate, analyze, and visualize data. The pretrial data pipeline requires improvements for all of these functions.

A **modular** application is one that can be relatively easily customized to plug into existing legacy systems, rather than requiring complex integrations or costly large-scale upgrades to the legacy system.

Modular data applications offer the following benefits:

- ▶ Data output is provided in a unified format for statewide aggregation and analysis.
- ▶ New functionality can be added rapidly (in days and weeks vs. months and years) and software bugs can be resolved efficiently.
- ▶ The applications are engineered to be scalable and customizable for integration with diverse existing systems.

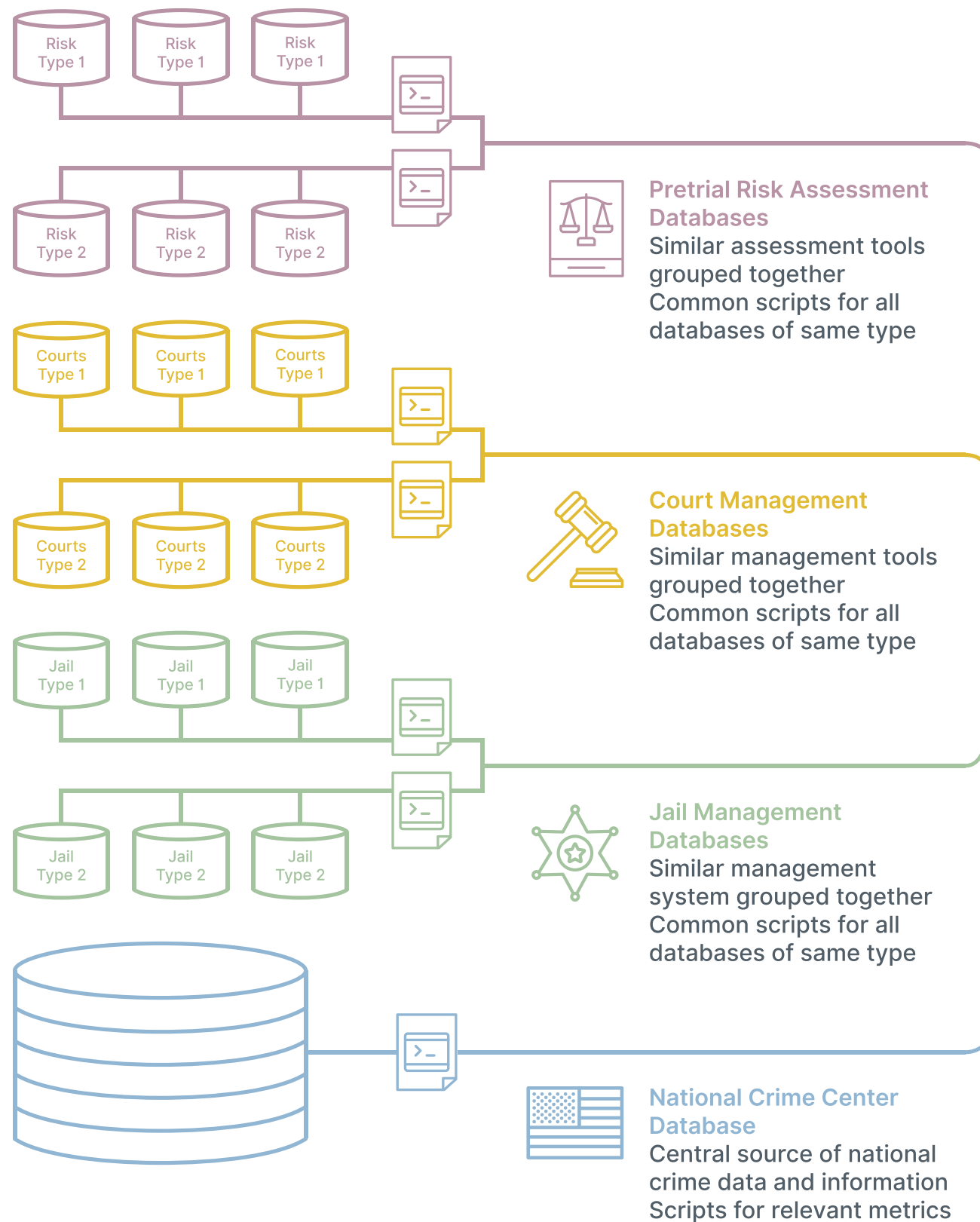
We recommend building data applications for the functions described above and deploying them across counties. By approaching county pretrial and jail management systems as groups of similar systems, the Judicial Council will be able to reduce the technical overhead and debt associated with viewing each county as its own system. Technology affords the greatest reductions in cost when implemented at scale, and this approach will help unify systems and data streams across multiple counties.

Integrating these data pipelines as a separate layer, rather than replacing or upgrading existing systems, will also enable significantly faster development and deployment at a reduced cost. As new case management and data systems are brought online or discarded, the data pipelines can be easily adapted, upgraded, and scaled to accommodate new functionality.

The Recommended System diagram illustrates the approach for implementation described above.



# Recommended California Pretrial Data Pipeline



## KEY TAKEAWAYS

1. Achieve scale, reduce costs, and accelerate development by taking a system-centric approach to pipeline development
2. Reduce the need to build custom integrations for each county by deploying solutions once for similar systems

## OVERVIEW OF RECOMMENDED CALIFORNIA PRETRIAL DATA ECOSYSTEM

Technical strategy is built on achieving scale by **developing common data pipeline solutions** for similar software systems

**New data pipelines can be built on top of existing systems** with a reduced need to “upgrade” an entire platform state-wide prior to beginning work

Counties and the JC **share and reuse code by using Open Source Software (OSS)** and modern software development practices

Counties and the Judicial Council **hire the “right” engineering personnel** to work together to develop solutions together and share knowledge

## ADVANTAGES

Lower investment in technology by each county and the JC by building and deploying solutions only once for similar software systems

Counties **benefit from shared solutions** and resources

Achieve significant **cost benefits** through reuse of code and modular development

Accelerated development by engaging the OSS community and working with contractors that use **modern software development practices**

**New features can be added and scaled in days and weeks** as opposed to months and years

Additional cost savings achieved if IT infrastructure is cloud-based (vs. local data centers)

## DISADVANTAGES

**Burden to lead rests with Judicial Council** including setting the strategy and hiring modern software engineers and contractors

Limited consolidation across multiple systems still **introduces technical and management complexity**

**NOTE:** This diagram depicts a **simplified representation of this solution**, and does not encompass every step and entity relevant to this process

## Hire software engineers experienced in building data pipelines (not CTOs)

We spoke with representatives from five California county pretrial departments who universally expressed a need for more technical personnel. Rather than adding Chief Technology Officer (CTO)-level personnel, the Judicial Council and local jurisdictions should opt to hire software data engineers, technical data analysts, and developer operations (DevOps) engineers. Below, we outline recommendations for hiring, as well as job descriptions that can be used to hire staff capable of building the recommended applications and pipelines.

- ▶ **Hire internally.** The state should follow in the footsteps of successful federal government programs like the [United States Digital Service](#) (USDS) and 18F and provide incentives to engineers to apply their tech industry experience in service of a public mission. These hires are likely to be more valuable to a team, and likely less expensive, than an outside consultant or contractor.
- ▶ **Hire self-starters who think outside of the box.** The state should hire adaptable engineers who can learn a diverse set of legacy systems and programming languages, and reverse-engineer antiquated systems when necessary.
- ▶ **Hire experts in software development and deployment (not CTOs).** Data pipelines and pretrial information systems should be regularly iterated upon, in accordance with standard software industry practice. Iteration should include new feature development, security updates, and the writing of scripts that automate the data pipeline as much as possible. Experience has shown that those best suited to iteration tend to be engineers who are familiar with modern software development practices such as agile development and who are comfortable using open-source tools. The Judicial Council and local counties should prioritize hiring engineers who are able to build the required data pipelines and avoid adding management layers such as CTOs.
- ▶ **Hire DevOps professionals.** “Developer Operations” or DevOps is a highly important function that enables the principles of continuous integration and continuous deployment (CI/CD), a critical capability for minimizing service disruption when implementing bug fixes, deploying new functionality, and improving system reliability. Developer Operations is an area of engineering that combines developers’ needs with IT operations. In contrast to traditional enterprise software development models, CI/CD allows for new features, bug fixes and design updates to be available as soon as the code is completed. Code can be deployed multiple times a day, instead of every six months or every year, as is common among more traditional enterprise software development companies. On the other hand, companies like Facebook, Amazon, and Apple—as well as many modern smaller development companies—upgrade their software hundreds or even thousands of times per day. DevOps engineers are experienced in the highly specialized field of building CI/CD pipelines.



### SAMPLE JOB DESCRIPTION

#### Data Engineer

A back-end software engineer who builds and integrates data pipelines from various resources to external stakeholders. Works closely with data analysts and domain experts in Law Enforcement, Probation, and Pretrial to automate data pipelines for extraction, transformation, cleaning, and testing.

#### Responsibilities

- ▶ Work closely with Data Analysts to specify data formats and expected behaviors for data pipeline outputs;
- ▶ Write scripts to extract data from legacy systems and transform that data into modern data formats;
- ▶ Build testing suites supporting data transformation, data structures, metadata, and dependency management over time;
- ▶ Use modern version control and continuous integration processes for data and code management;
- ▶ Maintain a unit testing framework and alerts for workflow failures;
- ▶ Repair and update data workflows given stakeholder input; and
- ▶ Collaborate closely with IT personnel at the state and local level.

#### Qualifications

- ▶ Industry, academic, or government experience in a similar data engineering or back-end software development role;
- ▶ Experience deploying and monitoring container platforms using technologies such as Docker;
- ▶ Experience processing and joining information between large disconnected datasets;
- ▶ Effective communication and experience writing clear documentation for teams with diverse skill sets; and
- ▶ Proficient experience in back-end software engineering and/or data engineering tasks.

Compensation: \$110–130k/year



**SAMPLE JOB DESCRIPTION**

**Developer Operations (DevOps) Engineer**

A DevOps engineer who can manage live data systems by quickly responding to new feature requests, bug fixes, customization requests, or use case expansions, helping data systems to scale across jurisdictions and pretrial agencies and continue to function effectively and efficiently.

Responsibilities

- ▶ Apply knowledge of cloud computing to automate the deployment of software, upgrades, and fixes;
- ▶ Declare server environments in code (infrastructure as code);
- ▶ Implement automation tools (CI/CD pipelines); and
- ▶ Conduct systems tests for security, performance, scalability, and availability.

Qualifications

- ▶ Bachelor’s degree in Computer Science, Engineering, or a relevant field;
- ▶ Expertise in code deployment tools (Docker, Puppet, Ansible, and Chef); and
- ▶ Strong command of software-automation production systems (Jenkins and Selenium).

Compensation: \$130-150k/year



**SAMPLE JOB DESCRIPTION**

**Data Analyst**

A data analyst who writes code to verify the correctness of the metrics collected from the various pretrial systems around the state. Designs the data pipeline cleaning, merging, metrics, and testing processes for data aggregation and presentation.

Responsibilities

- ▶ Work closely with the Data Engineer, IT departments, and county stakeholders to craft descriptive statistics and dashboards;
- ▶ Clean, merge, and analyze data from justice system partners; and
- ▶ Provide quarterly interactive reports.

Qualifications

- ▶ Experience in data analysis with Python or R;
- ▶ Experience using dashboarding software such as PowerBI or Tableau;
- ▶ Domain knowledge of the justice system not required, but preferred; and
- ▶ Effective communication and experience writing clear documentation for readers with diverse skill sets.

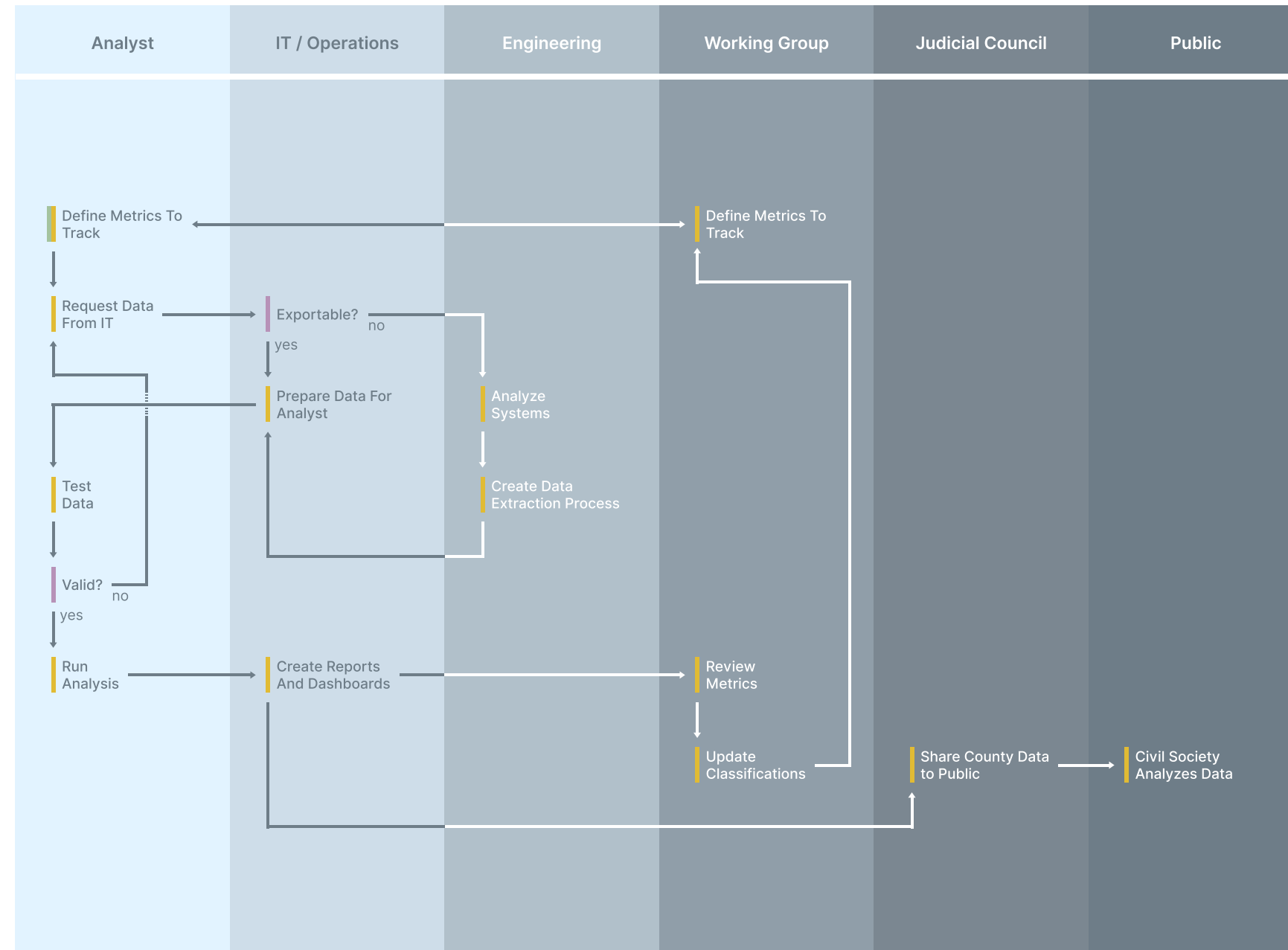
Compensation: \$110-130k/year

The Swim Lane Diagram on page 22 shows an example of the roles and responsibilities that technical staff typically hold in a modern data pipeline process. Hiring the technical personnel described above and instituting a process similar to that shown in the diagram will aid counties and the Judicial Council in the extraction, analysis, and visualization of data.

**Work with modern software development companies using open-source software**

In addition (or as an alternative) to hiring technical experts directly, we recommend that the Judicial Council work with mission-driven modern software companies to build out the modular software applications described in the first recommendation. We describe the difference between traditional enterprise software contractors and smaller modern software development companies below. We also provide examples of three companies currently working in the criminal justice space. We also highlight Santa Clara County’s pretrial data pipeline, which was built in-house.

## Swim Lanes: Roles and Responsibilities



Start

Action

Decision

### Enterprise Contracting vs. Small, Modern Software Companies

Federal and state governments typically choose to contract with large-to-medium-sized companies that have experience building and supporting enterprise-level software and also understand how to navigate complex government procurement processes. A typical IT contract usually involves developing detailed product requirements with the customer; developing the software (or customizing existing software) over an extended period of time; testing the software in accordance with a risk-management framework; and deploying it in a staged and generally highly disruptive process. The software development process typically takes between three and five years. Costs vary, but contracts of between \$10–\$200 million are not unusual. These contracts usually take longer than expected, endure expensive cost overruns, and deliver software that does not do what it is supposed to.<sup>12</sup>

By comparison, startups build and deploy software with much lower initial funding, yet successfully create, launch, and scale systems over time through iteration. Startups begin with seed amounts in the low \$100,000s; design and build software with rigorous user research; test software as it is being developed; deploy initial features in weeks or months; update iteratively and continuously; and launch entirely new applications within months that can scale to serve millions of users effectively and safely because they rely upon user-centered design and ongoing iteration. Because it is widely available and deployable, using open-source software (OSS) reduces costs even further and accelerates delivery, while offering transparency and high security standards.

### Examples of Effective Alternatives to Enterprise Contractors

We have collected several examples of approaches that can achieve criminal justice data reform goals efficiently and effectively. Recidiviz and OpenLattice are two organizations that have built successful solutions for criminal justice-related data pipelines using OSS and an iterative approach. Santa Clara County’s POPS system is an example of a county-built modern system that follows contemporary agile principles to build technical tools and services for the county pretrial agency. California Policy Lab is an academic organization that maintains an effective partnership with San Francisco County. Additional details about each organization are presented below.

Recidiviz (<https://www.recidiviz.org>) is a non-profit organization that has built an “extensible platform designed to power data-driven interventions in criminal justice.” Their tools include validation tools for instruments; metrics, analytics and reports; and program evaluation tools. Their work is live in five states and 1000 counties<sup>13</sup> and serves 10 percent of the incarcerated population in the United States. Their platform is open source and extensible. We compared the data schema published by

the Judicial Council with Recidiviz’s software and confirmed that their system could be adapted to meet state requirements in a short amount of time.

OpenLattice (<https://openlattice.com/>) is another example of a company that has specifically built a platform for modern data infrastructure for social services, healthcare, and public safety. OpenLattice has a pretrial case management application that runs the Public Safety Assessment (PSA) instrument. The system is populated with data from statewide and local courts as well as from law enforcement agencies. Similar to Recidiviz, OpenLattice’s software is cloud-based, published on GitHub, and open source.

A third option is the Pretrial Online Production System (POPS) built by Santa Clara County. Their tool was developed by engineers at the County and is regarded as a success among many counties in California.<sup>14</sup> One downside to POPS is that has not been released as open source. However, the Judicial Council could partner with Santa Clara County to publish the software as open source and share it with other counties.

Lastly, the California Policy Lab (<https://www.capolicylab.org>), a non-profit organization partnered with UC Berkeley and UCLA, has worked with the San Francisco Pretrial and the Sheriff’s Office to develop a dashboard for pretrial risk assessment that is used by the San Francisco pretrial working group to track outcomes on a quarterly basis. As an academic research institution, California Policy Lab serves as an independent voice to help track outcomes and provide guidance on policy to local jurisdictions and statewide.

### Use open-source software by default and as a matter of policy

We recommend using open-source software (OSS) to build the software applications and data pipelines described above, and to work with companies that create open-source software as a matter of practice. We also recommend that the Judicial Council implement a policy requiring software development to “default to open source” for all new systems, whether they are developed in-house, procured through vendors and contractors, or custom developed with enterprise software companies.

The value of open-source software is that it is available for free; however, this also fuels the misconception that OSS code is poorly written or is not suitable to handle confidential information. In fact, the opposite is true. OSS is developed openly and transparently by a community of developers, which helps ensure the code’s integrity. Another misperception is that the data used by OSS is openly shared; however, transparency of code in no way means that the data feeding into the software is freely accessible.

To ensure data security in complement with using open-source software, organizations should apply the latest data standards, such as the Advanced Encryp-

tion Standard (AES), to require that all data be encrypted in transit (when being shared) and in rest (when stored). Combining OSS with encryption standards for data ensures that similar systems can share the same code, significantly reducing redundancy in development and the cost of implementing technology for the state, while also ensuring data remains private.

The [Federal Source Code Policy](#) presents an excellent resource based on the federal government’s approach to “achieving efficiency, transparency, and innovation through reusable and open-source software.”<sup>15</sup> While the government has fallen short of its own stated goals in practice, the policy provides significant resources and rationale for using OSS and modern security practices. 18F, a digital services delivery agency of the General Services Administration (GSA), defaults to OSS on all its projects as a condition of engagement.<sup>16</sup>

The U.S. Digital Service, a federal agency established by the Obama Administration after the unsuccessful launch of Healthcare.gov, has developed its own playbook on best practices in software development, open source, and digital services.<sup>17</sup>

## Next Steps

The Judicial Council should allocate the remaining 10 percent of the appropriated \$75 million in funding to implement the recommendations listed above. Next steps include:

- ▶ Partner with the California Chief Technology Officer (CTO) to work together on the technology recommendations made in this paper.
- ▶ Hire a small technical team to begin the process of analyzing each county’s data pipeline needs, as recommended.
- ▶ Begin meeting with representatives from counties to understand their needs. Survey county data systems and group identical or similar systems for shared data extraction tooling. This process should take no more than 3–4 months.
- ▶ Standardize reporting around a common set of desired metrics and outcomes to meet the Judicial Council’s reporting needs, to frequently validate the pretrial instrument being used, and to prepare to meet the requirements of SB–36.
- ▶ Develop state-level requirements that present critical metrics and analyses across jurisdictions and in aggregate. Define a cadence to update the outcome information frequently (at least every 24 hours).
- ▶ Develop county-based requirements for outcomes and ensure that each county and the Judicial Council are working toward similar goals.

Ensure that each county has an independent pretrial working group that meets regularly and is exclusively focused on data collection, sharing, validation, and feature development.

## Who We Are

[The Aspen Tech Policy Hub](#), part of the [Aspen Institute](#), is a West Coast policy incubator, training a new generation of tech policy entrepreneurs. Modeled after tech incubators like Y Combinator, we take tech experts, teach them the policy process through an in-residence fellowship program in the Bay Area, and encourage them to develop outside-the-box solutions to society’s problems.

**Allison Day** is working at the cutting edge of artificial intelligence, coordinating engineering and natural language processing (NLP) teams to build the Google Assistant. She also has experience on the front lines of gender-based violence response and is building tech for humanitarian relief at the border. She holds an MA in Theoretical Linguistics from the University of California at Santa Cruz. Allison is a polyglot, a performer, and an activist striving to unite and inspire people towards change and equity.

**Anil Dewan** is a Product Manager with the Defense Digital Service (DDS) at the Department of Defense. He was previously a co-founder of Pollen Labs, a location-based community and social media start up, and Director of Interactive Media at KCRW, the NPR station and culture hub in Los Angeles. Anil has an MS in Systems and Safety Management from the University of Southern California. In his spare time, Anil likes to research the implications of technology and AI on our lives and their potential impact on our future.

**Karissa McKelvey** is a product and engineering professional whose work has been depended upon by at-risk users—including journalists, human rights defenders, and civil society activists—who live within repressive environments and wish to speak freely online. Karissa has delivered public-centered initiatives for open-source projects, non-profit organizations, and startups that leverage emerging technologies. In 2016, she co-founded Code for Science and Society to systematize positive outcomes across the ecosystem of open-source public interest technology.

## Endnotes

- 1 Joshua Aiken, “Era of Mass Expansion: Why State Officials Should Fight Jail Growth,” *Prison Policy Initiative*, <https://www.prisonpolicy.org/reports/jailovertime.html> (May 31, 2017).
- 2 Aiken, *supra* note 1.
- 3 Partnership on AI, *Report on Algorithmic Risk Assessment Tools in the U.S. Criminal Justice System*, <https://www.partnershiponai.org/report-on-machine-learning-in-risk-assessment-tools-in-the-u-s-criminal-justice-system/> (2019), 31.
- 4 Partnership on AI, *supra* note 3, at 3; John Logan Koepke & David G. Robinson, *Danger Ahead: Risk Assessment and the Future of Bail Reform*, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3041622](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3041622) (Upturn, updated December 31, 2018).
- 5 Megan Stevenson, “Assessing Risk Assessment in Action,” *Minnesota Law Review* 103, no. 303 (Feb 2019).
- 6 Sarah L. Desmarai and Evan M. Lowder, “Pre-Trial Risk Assessment Tools: A Primer for Judges, Prosecutors, and Defense Attorneys,” *MacArthur Foundation Safety and Justice Challenge*, <http://www.safetyandjusticechallenge.org/wp-content/uploads/2019/02/Pretrial-Risk-Assessment-Primer-February-2019.pdf> (February 2019).
- 7 Examples of metrics include release rates, rates of recidivism, failure to appear rates, judge concurrence with risk assessment recommendations, etc.
- 8 Pretrial Justice Institute, “Santa Clara County Case Study: Using Technology and Innovation to Improve Client Contact,” *Pretrial Justice Institute Case Studies* 3 no. 1 (January 2019).
- 9 Lisa Brooks, Meghan Guevara, and Sam Packard, Californians for Safety and Justice “Pretrial Progress: A Survey of Pretrial Practices and Services in California,” *Crime and Justice Institute at Community Resources for Justice* (August 2015).
- 10 Judicial Council of California, “Judicial Council Funds 16 Pretrial Pilot Programs,” *California Courts Newsroom*, <https://newsroom.courts.ca.gov/news/judicial-council-funds-16-pretrial-pilot-programs>, (August 9, 2019).
- 11 Judicial Council of California, “Pretrial Reform and Operations Workgroup: Pretrial Pilot Program Recommendations to the Judicial Council,” *California Courts Newsroom*, [https://newsroom.courts.ca.gov/internal\\_redirect/cms.ipressroom.com.s3.amazonaws.com/262/files/20197/Pretrial%20Pilot%20Programs%20fact%20sheet\\_201908091752.pdf](https://newsroom.courts.ca.gov/internal_redirect/cms.ipressroom.com.s3.amazonaws.com/262/files/20197/Pretrial%20Pilot%20Programs%20fact%20sheet_201908091752.pdf), (August 9, 2019); Judicial Council of California, “Pretrial Pilot Program Court Questions & Responses,” *California Courts Newsroom*, <https://www.courts.ca.gov/documents/pdr-pretrialpilotprograms-faqs.pdf>, (June 21, 2019).
- 12 Wes Venteicher, “California’s Failed Tech Projects Come In Late and Over Budget. Gavin Newsom Wants To Fix Them,” *Sacramento Bee*, <https://www.sacbee.com/news/politics-government/the-state-worker/article227206314.html> (March 7, 2019).
- 13 “Recidiviz: A Criminal Justice Data Platform,” Recidiviz, <https://www.recidiviz.org/> (accessed November 2019).
- 14 Pretrial Justice Institute, *supra* note 8.
- 15 Tony Scott and Anne E. Rung, “Federal Source Code Policy: Achieving Efficiency, Transparency, and Innovation through Reusable and Open Source Software,” *Federal Source Code Policy*, <https://source-code.cio.gov> (accessed November 2019).
- 16 Raphael Majma and Eric Mill, “18F: An Open Source Team,” 18F Blog, <https://18f.gsa.gov/2014/07/29/18f-an-open-source-team> (July 29, 2014); 18F, General Services Administration, “18F: An Open Source Team,” 18F *Open Source Policy*, <https://github.com/18F/open-source-policy/blob/master/policy.md> (February 21, 2018).
- 17 United States Digital Service, “*Digital Services Playbook*,” <https://playbook.cio.gov/> (accessed November 2019).



**ASPEN TECH  
POLICY HUB**

 THE ASPEN INSTITUTE