Fall 2020

System Change Assessment and Support: Approaches from the Data Driven Recovery Project

Contract # 19MHSOAC074, Deliverable 1

Submitted to: Mental Health Services Oversight and Accountability Commission

Kevin O'Connell

KEVIN@OCONNELLRESEARCH.COM

TABLE OF CONTENTS

This document lays out the key elements of the Data Driven Recovery Project, or DDRP. The goal is to both present the vision of DDRP, as well as some tangible approaches to localizing data driven efforts to create a system of care that meets the goals of equitable justice, public safety, and recovery for those with behavioral health needs. Counties are often faced with the need to make quick decisions with the best information available. When looking across systems at some of their community's most pressing needs, having a strategy as well as the data to answer various questions is imperative.

Recovery for any one person is not just about having the right services but having the right systems that effectively coordinate the correct level of resources and care at the right moments. As a direct service technical assistance initiative, the project is focused on county level capacity-building to create ongoing approaches and tools, and lay the groundwork for system innovation, as well as programmatic innovation. DDRP is focused on delivering 4 areas of work:

- Applying timely analysis, as well as the creation of baselines
- Development of system maps, program inventories and shared priorities
- Create a local data governance strategy to sustain analysis
- Develop an integrated database across justice, behavioral health, and social services

DDRP's core mission is to provide data and information to county staff and leaders to better inform their choices about how they use local initiatives, compared to those from other communities. By looking at ways to both understand and innovate based on a shared understanding of the clients in multiple systems of care, data driven practices become more engrained as way of bringing the best available data to bear in improving outcomes for clients. Data is often the start of a conversation, not the answer to questions.

MEET COUNTIES WHERE THEY ARE

The Data Driven Recovery Project (DDRP) was started with the aim to "meet counties where they are". That included both their current technical capacity as well as the issues that counties were attempting to address regarding the improvement of outcomes for behavioral health clients. Ten counties have opted to participate in the DDRP to look at system level opportunities for collaboration, but also to be specific about shared impacts and goals. The support offered to counties is focused on developing localized plans and priorities, a county-level strategy combining data integration, and an ongoing structure to use it strategically. The DDRP focuses on capacity building, both in building technical infrastructure to integrate administrative data across agencies, as well as championing the ongoing use of analysis and data-informed practices. Appendix 1 has common list of data fields, as well as link to example datasets.

Complex legal, correctional, treatment, and social service delivery requires a better understanding of each system's constraints, funding, and desired outcomes in the long term. The ever-evolving relationship with the state gives DDRP a unique perspective in helping counties navigate local realities, as well as connect to larger themes and policy shifts occurring in numerous areas of justice and human services. Appendix 2 presents a basic approach for working through integrating data across systems.

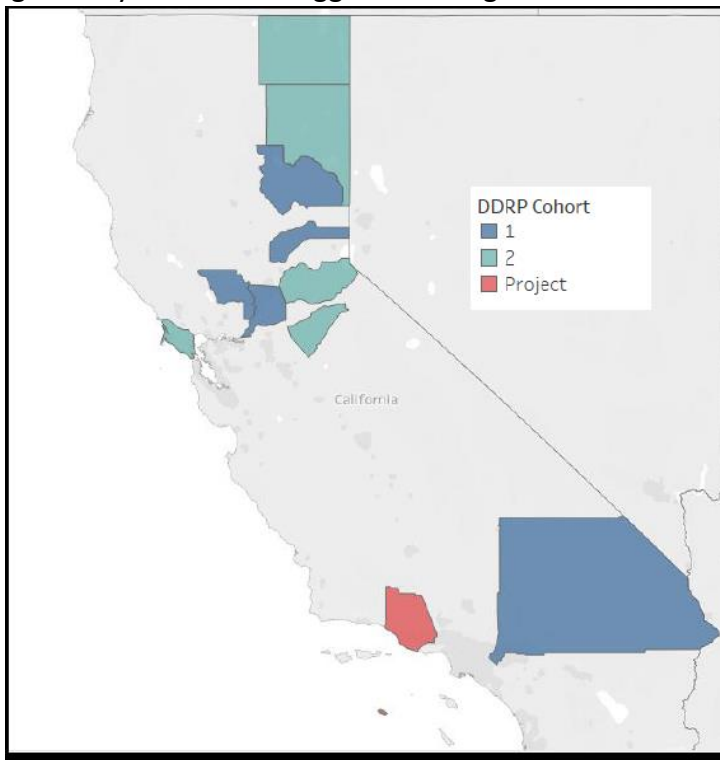## BIG IMPROVEMENTS CAN COME IN INCREMENTAL CHANGES

A barrier to innovation is often the myth that changes take lots of money or trying something that has not been done before. Looking at aggregated data across systems and time can reveal surprises and challenge assumptions. Utilizing a philosophy of humble inquiry can lead to new thinking and new ideas, and innovations. Changes in one part of the justice behavioral health continuum can often present an opportunity in another, and what can follow is a more integrated approach between state and local priorities. There are many conduits for communicating needs, ideas, and best practices, but too often they leave counties without dedicated resources to connect the dots and maintain focus on system issues.

DDRP's vision is to provide a useful framework for sharing the effective use of data not just within counties, but between them as well. Creating these pathways can help even the playing field when there are technical gaps, or when it takes a reference point from another county to better understand localizing an innovation. In the end, there should be value placed on incremental improvement that is specific and actionable, rather than dependence on outside entities. The needs will always be far greater than the resources, so data should not act alone, but be paired with values and shared vision.

Strategies to make this happen can be grounded in state initiatives as well as active grant seeking, but they should still consistently refer to values and alignment. DDRP comes in not as strategy in and of itself, but rather as a laboratory for exploring and acting locally. It is then possible to find alignment with state and federal actions.

COLLABORATIVE DATA UTILIZATION SHOULD LEAD THE WAY

By emphasizing areas of shared importance for individual counties, the goal is to have the maximum impact on people's outcomes in treatment and avoid incarceration or court involvement that may be confounded by behavioral health needs. Through development of workplan that reflects multiple goals, but with a vision for sustainability and demonstrating utility, the ability to overcome local barriers starts to grow. Some of the biggest barriers to integrating data is a range of political, legal and framing issues, but the technical pieces generally are not the biggest challenge. This is not to understate the technical complexity of data exchanges or complex IT projects, but more to highlight that using data to drive innovation and system improvement can be done by starting with some basic ingredients and goals. Counties often are forced to use informal means to connect across agencies, as practitioners know that their client's lives overlap with multiple services and systems.

The tradeoff for efficiency should never be quality, as "doing more with less" is not a viable solution for reducing justice involvement and promoting recovery and stability. Leaning too heavily on evidence-based solutions crowds out local innovation that might be more cost effective for the clients who are best served by avoiding deeper justice system involvement.

Using data to have a "client centered" view of the justice process and treatment pathways does not always cost more and can leverage more parts of a local team connected to someone's care. That said, the project is also meant to be a bridge to create evidence locally, not only in

raw numbers or impacts, but by helping to understand why something works across justice and behavioral health.

Data is only as useful as what can be done with it. A core strategy of DDRP is to leverage an integrated database of people in both justice and behavioral health to get a system-level view of a county, then to zoom in to specific opportunities, challenges, and areas of deeper exploration.

This analysis is then used by local workgroups, community forums, or the Board of Supervisors to better inform system choices. By building system and program maps in conjunction, the county then has a shared sense of the current system, as well as how it is being used. This gives counties a roadmap for sustaining data integration, as well guidelines for strategies to obtain grant money, reorient systems to resources, and better partner when looking to evaluate "what works". Numerous state initiatives also look to counties to inform the state narrative, however this is often built on specific grants or programs rather than on a full understanding of the universe of people touching the justice system and needing behavioral health care. Appendix 3 lists an approach for develop workgroups and structure around how data might be used for multiple analytic purposes.

*Phase 2
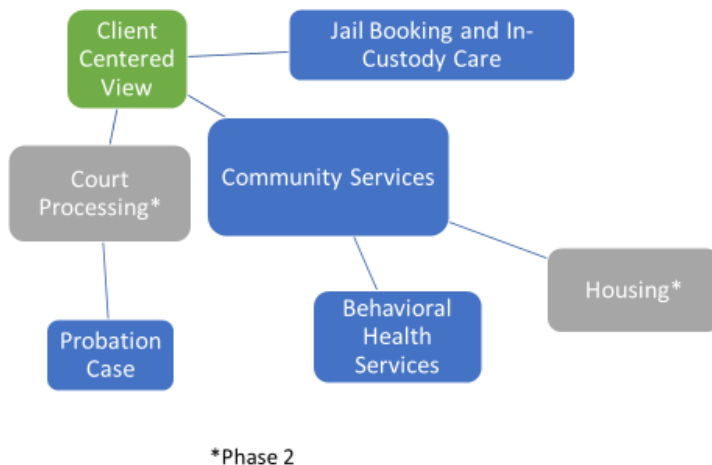
DDRP's unique setup to offer technical assistance to counties directly, via a county consortium led by Yolo County, allows the project to operate similarly to an embedded research team rather than a new external initiative.   The core set of tools, approaches, materials, and conversations are carried forward to a monthly community of practice call. The call follows a format of dialogue as well as presentations on areas of shared interest.

The goal of DDRP is to help communities develop an understanding of general client trajectories based on patterns and then develop system changes, programmatic innovations, or connections to improve clients lives. Although the project is state funded, all agreements, data and data sharing approaches remain with the county and the local data owner. A unified data and analysis strategy help bring consistency to the results, mirroring how a county might merge
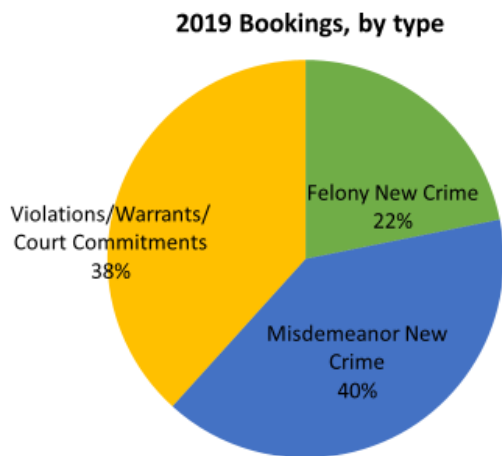
and use integrated data locally. In the end, outside expertise only goes so far, but offering the tools and approaches for local innovation can help embed approaches for the long term.

Counties start by assessing their justice and behavioral health systems for who enters what systems, how long they stay, and who comes back. This aggregate data analysis gives counties a better idea of the drivers of justice involvement, as well as a base understanding of the prevalence of behavioral health clients in various parts of the justice system. Leveraging integrated data allows counties to then look at specific areas that drive disproportionate justice involvement or treatment engagement.

This initial overview is meant to give stakeholders a clear set of facts to then develop more specific policy proposals and discussions. What follows is a generalized look at work in various DDRP counties as well as to show how different counties have leveraged their data to better understand their client's movements in various systems.

Understanding a jail's use is an essential starting point to better understand the range of bookings and daily population to avoid bookings, divert them, or create interventions. These are important differences as too often the justice-involved population is grouped together, instead of exploring different specific policies and ideas to assist them. This lack of specificity leads to solutions that are under-resourced or mismatched to the type of person in jail. Jails are complex places, and often those in jail are there for things other than new crime. The figure shows an example DDRP county where 38% of those entering jail are there for things like warrants, supervision violations, or commitments from court. In this example site, felony new crime bookings make up 22%. There are several options a county can take just with this data regarding how people enter jail, since the options and court process for new crimes is different than those for non-new crimes.

2019 Bookings, by type

- Felony New Crime 22%
- Misdemeanor New Crime 40%
- Violations/Warrants/Court Commitments 38%

The admissions into jail look quite different than those in custody on a given day. In most jails, most people booked are released within four days of entry, pointing to strategies for meeting needs of a population that enters and leaves jail very quickly. When overlaid with mental health screening data in the figure to the right, the amount of people with possible needs for mental health is more nuanced as it shows 60% of people are released in a few days, but with a significant amount of BH need (nearly 30%). Jail mental health screening tools are used as a brief assessment to then trigger further treatment and follow up for in-custody care, but they can also give communities a better sense of how people with possible mental health needs are moving through their system.
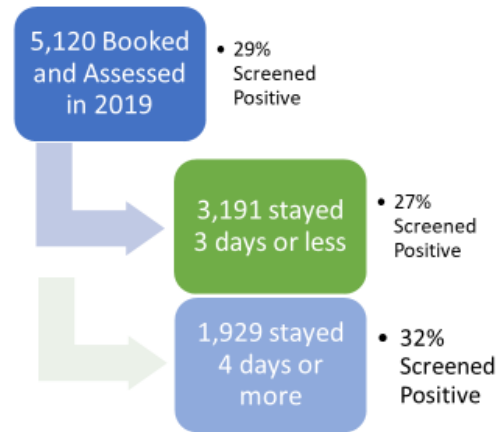
Using these same assessment tools can then be useful to see where mental health issues are prevalent in terms of how people entered jail. The figure below shows the volume of people booked into jail for felonies, overlaid with the number of people screening positive for mental needs. With the booking reasons broken out by how a person entered jail[1], analysis can focus on those in jail for similar reasons and look to more specific ideas or
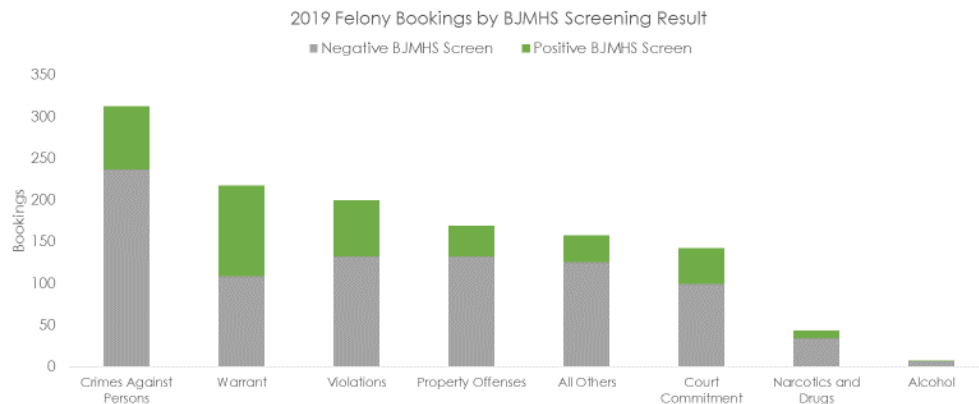
solutions when those with mental health needs are overrepresented. Are there ways to avoid warrant bookings by reducing failures to appear in court for those with mental health issues? Can violations of supervision be reduced by looking at alternative sanction matrices or partnering with clinicians? Do certain types of law violations or booking reasons offer different options when looked at in aggregate?

---

[1] Crime statutes are mapped to DOJ summary codes, then within a booking only the most serious charge retained using the DOJ's seriousness hierarchy. Booking and entry types are then prioritized for those representing new crimes versus other entries such as supervision violations or

Misdemeanor analysis offers a similar opportunity, often with much higher volumes of people. With misdemeanors, there is more direct evidence of behavioral health needs, as alcohol, drugs, and property crime often coincide with high need individuals.

2019 Misdemeanor Bookings by BJMH Screening Result
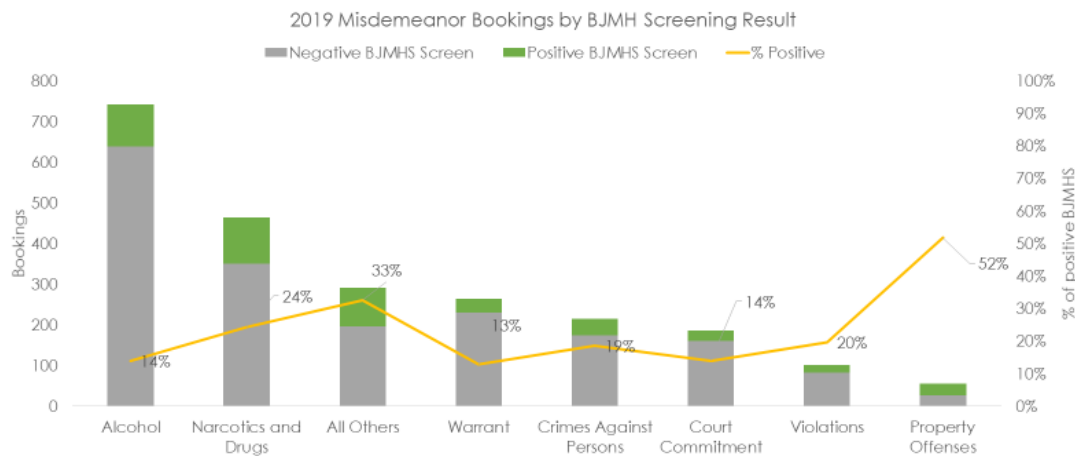
Looking at bookings helps to understand who enters jail, but there also can be disparities for those staying longer with mental health issues. Since so much of a jail's population leaves quickly, a more useful metric is examining those that stay more than 4 days. This can show differences in prevalence that then starts to drive disparities in average daily population. The figure below shows a slightly longer length of stay for those with mental health needs for new crimes, but a substantially longer length of stay for those entering for things like warrants, supervision violations, and court commitments. Why people stay can be due to the court process and pretrial release decisions, challenges in getting people into residential placements, or a range of other things.
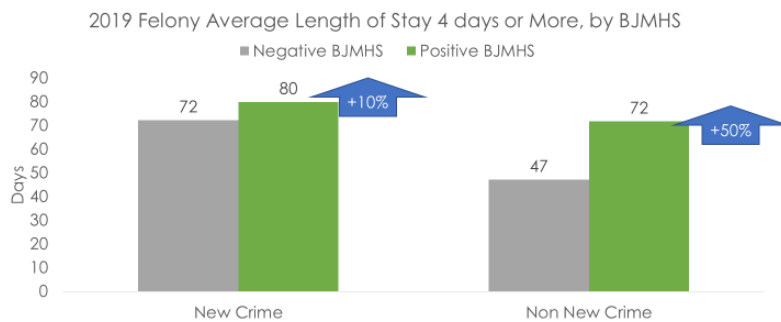
2019 Felony Average Length of Stay 4 days or More, by BJMHS

Regardless of the specific reason, all this needs to be explored and analyzed as there will not be one solution or explanation.

This same data can also help to show who is entering jail the most often and create opportunities for reducing the recurrence of those entering jail repeatedly. The figure below shows that 5% of the people booked into jail generate 20% of the booking activity. Examining these 130 people reveals high rates of mental health need as well as low level crimes. This presents a good opportunity for social service solutions for those who are a low public safety risk but have high needs.
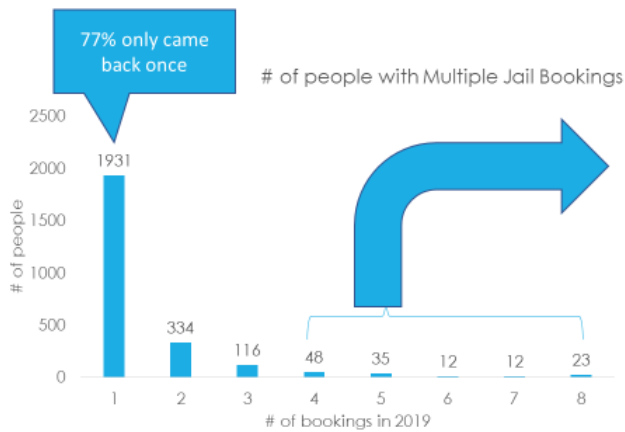
- 5% of people generated 20% of bookings

- 130 people came in more than 4 times, generated 806 bookings
  - 55% screened positive

- 75% Misdemeanors (2019 bookings)

System can then be overlaid with those receiving services in the community from behavioral health and entering custody to better understand service connections, revealing opportunities to better understand how services and justice trajectories overlap. The figure below can help behavioral health departments to more clearly identify where service delivery is often connected to justice interactions. For example, the client example below has a history of jail bookings for supervision violations and has sporadically accessed social services in the community. This approach can be used in case file review and quality assurance to better understand why people don't make connections to treatment at release. This wider focus allows for recognition of clients who will need connections to service at release, the creation of defined and data-driven client pathways, and` an understanding of how the client might experience these connections.
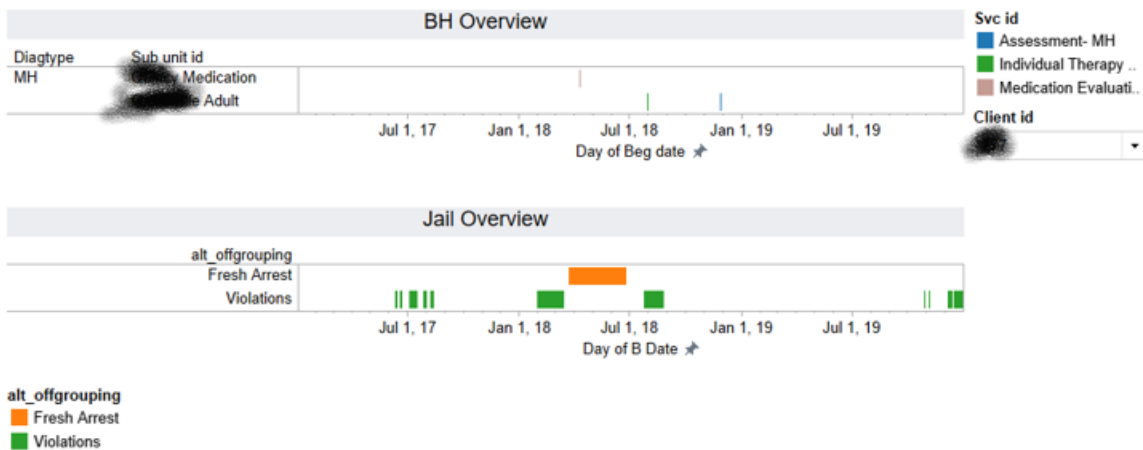
Linking this same treatment data with the probation population can also give an idea of how treatment needs overlap with those being supervised in the community. Probation, the most common sanction for those convicted, is meant to ensure the client is fulfilling court orders while also

Static Risk Assessments for those with MH Diagnosis (2019)

Risk Level	%
Low	12%
Low/Moderate	4%
Moderate	33%
High	40%
Very High	11%

50% are High Static Risk
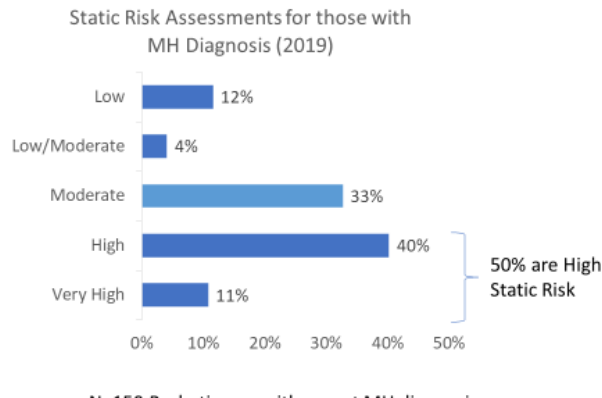
Those at higher risk for recidivism also had high needs in:

Need	% with a High Need in
Criminal Attitudes and Behavioral Patterns	6%
Criminal History	13%
Education, Employment, Financial	43%
Family and Social Support	19%
Neighborhood Context	64%
Peers Associations	21%
Substance Use	29%

facilitating a connection to services. By looking at those on probation with a recent behavioral health diagnosis, a county can better align treatment needs with criminogenic needs. For example, in the figure above, 150 people on probation also had recent assessments or interactions with behavioral health. 80% of those same clients were of medium or high risk to re-offend. Those clients of higher risk had a wide array of needs, of which probation would prioritize its resources to meet those to avoid future recidivism.

The result of such discussions and analysis are policy recommendations, founded in data and supported by a broad consensus of what is possible now to solve issues, as well as what could be possible after broader discussions and planning. Further analysis of the same data leads to better program scoping, scaling, and targeting, as well as evaluation. Using consistent baselines can help leaders focus on key metrics of overall success, but then create more specific monitoring of programs or operational practices.

Counties can create baselines that compare, at a high level, metrics that measure how strategies are reducing the number of people entering jail with mental health needs, reducing how long they stay, and hopefully reducing patterns of jail recurrence.   These types of metrics are a start, with many more possible to help analyze operations, as well as interpret things that matter for their programs and systems. The important piece is to choose metrics that

Numbers of Total Bookings and Unique Individuals Identified as screening positive for MH needs in 2019	1,158 people 1,504 bookings	2,848 people 3,619 bookings
Average Length of Stay in Jail for People screening positive for MH needs in 2019	16 days	14 days
3-year Jail Re-booking amounts for people screening positive for MH needs(2017 release cohort)	7 bookings	4 bookings

are meaningful and actionable, and that are attainable given available data.

By aligning priorities with strategies designed to impact key metrics, the county can align resources and programs, as well as highlight their evaluation funds. Although any change in outcomes cannot be assured to come from a specific intervention, aligning a planning process on the front end can prioritize discussions and focusing effort. The example in the figure to the right is from a county that created metrics, developed a program map, and uses its workgroup to align new ideas and strategies.

1. Keep people with SMI out of jail	2. Shorten length of stay in jail	3. Keep people from coming back to jail
• HOME team • Possible mobile crisis	• Work on ISTs • Mental Health Diversion	• Violations and warrants • Prop 47/CSIG grant and activities • Expanded SUD/MH services in jails

FACILITATE THE DEVELOPMENT OF SYSTEM MAPS, PROGRAM INVENTORIES AND SHARED PRIORITIES

Each county develops its own system map to assist in plotting justice system decision points to treatment pathways and opportunities for diversion. This is combined with a community stakeholder meeting, with local leaders leading the discussions and engagement about specific parts of the justice system.

Nevada County Sequential Intercept Map, Contact: Andrea Lehmkuhl (andrea.lehmkuhl@co.nevada.ca.us)
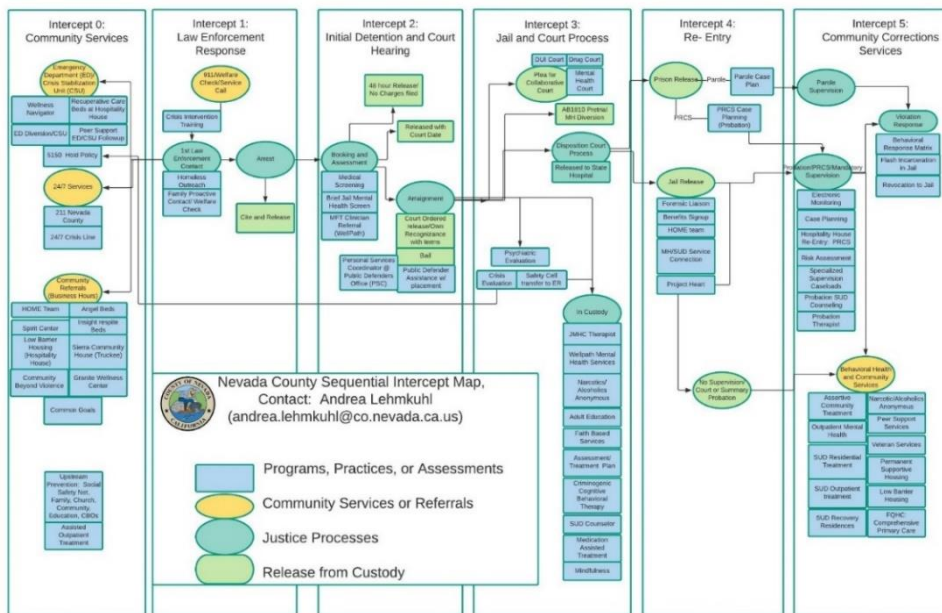
Meaningful cross-system collaboration is required to establish effective and efficient services for people with mental illness in the criminal justice system. The mapping process provides an opportunity for communities to visualize how behavioral health and criminal justice systems intersect in serving individuals with mental illnesses and co-occurring disorders. An example map from Nevada county in shows the range of investments and strategies.

Counties also use these maps to look at how they can employ data across the intercepts to measure how many people use what resources, as well as prevalence rates for different behavioral health needs, and then assign specific outcomes and metrics across the system.

The value in this is often that counties lack a single source of how many systems and programs impact clients.

By utilizing DDRP, agencies would share information about programs, strategies, and approaches in a constructive environment. Discussion in small groups is facilitated to then dive deeper into opportunities to:

- Refine and update existing program lists and services
- Discuss service gaps or needed program improvements
- Make recommendations on priorities

This results in a prioritized list with input from both the stakeholders and community that serves as a roadmap for analysis as well as policy action. It also gives community members a chance to give input on gaps, as well as hear from local system experts regarding how various

parts of the system work in more detail. The development of these priorities comes from the same stakeholder process that advances into a collaborative map of programs.
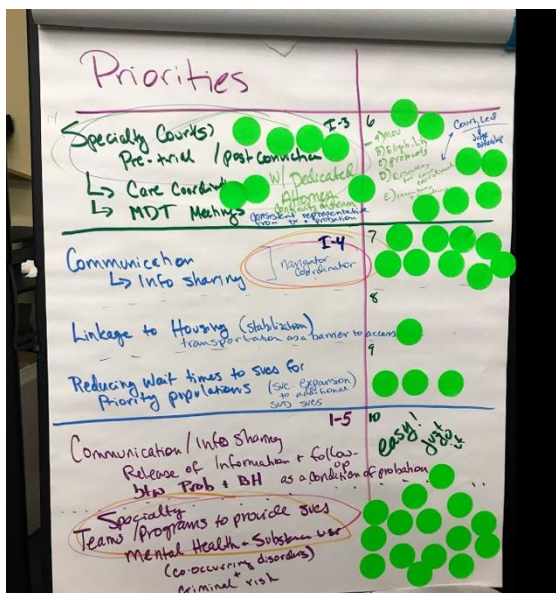
These system maps can then be the basis for a program inventory that gives participants another resource for exploring and understanding their county. The goal of a program inventory is to create a unified list of key information on programs, practices and assessments that includes information on delivery approaches, budgets, and evidence of effectiveness. The inventory approach also helps to give context to the goals and expected outcomes of specific interventions and allows a site to visualize how a series of interventions are connected.

CREATE A LOCAL DATA GOVERNANCE STRATEGY TO SUSTAIN ANALYSIS

A fundamental step is to create rules and processes for the development of a local database of people entering jail, on probation, going through the court process, moving through the crisis continuum, and engaging in treatment. The core strategy is to create rules and trust that allow data to be used for research and analysis while still maintaining client confidentiality.

Data governance is the guiding principle that manages the data assets and uses of data functions. Gathering and managing behavioral health and justice data in a single warehouse requires an improvement in infrastructure to guarantee availability, usability, integrity, and messaging. A data governance program includes a leadership group, a defined set of activities, and a plan to execute the procedures.

Data governance can be planned, managed, and implemented through a two-level structure ensuring a county-defined mix of executive level support and sponsorship, as well as subject matter experts. Strategic goals are to prioritize analysis efforts, communicate with or represent county data owners, and direct long term improvements in collection and integration.

Data governance also describes an evolutionary process, altering the county's way of thinking, and setting up the processes to handle information so that data may be utilized in more varied ways, such as:

- A representative of an established research organization or university interested in studying the County's behavioral health/justice-involved population and the effectiveness of various interventions the County employs
- A staff member of a County agency to learn more about the population that accesses the behavioral health and justice systems to inform how best to serve them
- Engaging a researcher serving as an external evaluator for a grant to make their time more efficient and useful, without having to navigate multiple agencies
- Using principles of "open data" to allow community members to identify the population they serve, and learn more about their characteristics to develop effective programs to address the population's needs

By starting with a clear sense of what is allowed and what is already part of a data warehouse, usage can be more rules-based, making it simpler, more secure, and respectful of client confidentiality in ways that inform and extend its use. The DDRP will be moving ahead with several counties in extending the use cases above.

Justice and behavioral health lack a common identifier in most counties. Matching allows a researcher to merge records through code, without passing any identifying information to other data owners. This database is then standardized and prepared to not just answer single questions, but many different questions. This strategy utilizes integration using a standardized coding scheme, making it efficient and reproducible. Creating the database requires some startup work, but once complete, updating it only involves re-running local queries. Using a federated data model as an approach for research and analysis lowers the level of effort to keep data updated, but ensures the data is only used for the purposes envisioned through the data governance process.

Both CORI and HIPPA offer exceptions for the use of Personally Identified Information (PII) for research and internal planning. Further, the use of the data is internal to the county, not a release of information to a third party. California Penal Code § 13202[2] grants analytic/research provisions as does HIPAA under 45 CFR 164.501 and 42 CFR § 2.52[3]. However, both require the removal of PII and avoidance of sharing information that might re-identify someone.

Using an intermediary step that merges PII to build a master list of people in both behavioral health and justice systems can then avoid explicitly sharing PII as merging and transformation would happen in a protected environment, and only using de-identified data for analysis. This merging of PII would occur using the phonetic algorithm, used to match names based on phonetic spelling of names, paired with the date of birth.[4]

---

[2] Notwithstanding subdivision (g) of Section 11105 and subdivision (a) of Section 13305 , every public agency or bona fide research body immediately concerned with the prevention or control of crime, the quality of criminal justice, or the custody or correction of offenders may be provided with such criminal offender record information as is required for the performance of its duties, provided that any material identifying individuals is not transferred, revealed, or used for other than research or statistical activities and reports or publications derived therefrom do not identify specific individuals, and provided that such agency or body pays the cost of the processing of such data as determined by the Attorney General.

[3] The HIPAA Privacy Rule establishes the conditions under which protected health information may be used or disclosed by covered entities for research purposes. Research is defined in the Privacy Rule as, "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge." See 45 CFR 164.501 and 42 CFR § 2.52. A covered entity may use or disclose for research purposes health information which has been de-identified (in accordance with 45 CFR 164.502(d), and 164.514(a)-(c) of the Rule) and 42 CFR § 2.52 (b)(3).

4 https://www.archives.gov/research/census/soundex.html

The merge of behavioral health and justice PII is then "pseudonymized[5]" to limit the exposure. Merged data is not passed back to the original data owner, so no new data or identifiers would be added to the original data owners' submissions or data flow. Pseudonymization does not remove all identifying information from the dataset, but merely reduces the clear relationship of a dataset with the original identity of an individual. The produced data warehouse resources would then be managed by workgroups/teams formed

Behavioral Health IDs of people and Events
- Social Security Number
- Name
- Date of Birth
- Client ID
- Health Agency ID, Episode or Referral ID

Justice IDs of people and Events
- Criminal Identifying Information (CII)
- Name
- Date of Birth
- Agency Person Identifiers (Jail, Probation, etc)
- Court Case or Docket Number
- Jail Booking Identifier
- Probation Referral Identifier

Figure 1: Example PII used in data merge

through the data governance process, working across agencies. The figure below shows how the initial loading of PII creates a master list of people common to all data owners, which is then merged with event and episode identifiers, after which PII is given a pseudonym so that the final dataset contains no PII.

Behavioral Health — Extract service and Episode fields — Run Soundex algorithm on PII

Jail Management System — Extract Booking and charge information

Run Soundex algorithm on PII

BH/CJ Master PII table, merged on soundex name

Run soundex algorithm on PII from Probation, add to CH/CJ

Pseudonymization of PII, creating personalized client IDs mapped to agency data

PII and Episode ID Pseudonymized

Unified CJ/BH database, without PII

Probation

In-custody Health Services

Extract case processing, referral, probation, and risk assessment data

Since each data owner has unique workflows and data structures, the initial data extraction would attempt to create the basic flow of people through each system over time, then carefully build out common baselines and analysis.

All five Phase-1 counties have expressed interest in this merge warehouse, with four completing this step in the first year. Counties in Cohort 2 will be starting this step during fall 2020, with a similar set of governance documents, data dictionaries, and code.

---

5 To pseudonymize a data set, the additional information must be kept separately and subject to technical and organizational measures to ensure non-attribution to an identified or identifiable person.

LOOKING AHEAD

As phase 2 begins, the DDRP is working to add 5 new counties, as well as launch more targeted, specific projects where counties can engage with some of the tools and approaches. Connections to other MHSOAC funded county innovation projects can also utilize the analytic approaches that inform system design, measurements, and improvements for clients served. The DDRP will also be integrating itself with other Innovation Incubator projects like CrisisNOW and Fiscal Sustainability where needed, as well as deepening connections on state level policy initiatives.

By offering analytics and data integration advice as a more adaptable service, it will reach more counties. This decentralized approach is based on keeping people's identifies confidential, such that DDRP develops relationships with each data owner. The trust of the project and the process becomes part of the roadmap for counties to share data for the purposes of research more readily, and model the technical governance approaches taken, long after a DDRP engagement is complete. The goal of capacity building is giving counties the tools and use cases to then choose the best path forward and provide county-specific assistance to inform local practice.

APPENDIX 1: GENERAL SPECIFICATIONS FOR DATA FIELDS

BEHAVIORAL HEALTH (COMMUNITY AND IN-CUSTODY)

Variable Name	Description
Behavioral Health ID	Behavioral Health Identifier
Episode ID	Unique Identifier of the intervention Episode
Service ID	Unique identifier of the service
Episode Start date	Start date of the episode
Episode End Date	End date of the episode
Priority Diagnosis	Priority Diagnosis during Episode
Treatment Setting	Type of setting for Episode
Service date	Date of Service
Service type	Service type (Assessment, Treatment, Case Management)
Service Duration	Duration of Service in minutes
Location	Location of Service

Example dataset for community Behavioral health (link)

Example data for In custody behavioral health (link)

ICD10 Codes with crosswalk to diagnosis grouping and type (link)

JAIL

Variable Name	Description
Inmate ID	Inmate ID in the jail System
Booking ID	Jail Booking Number
Gender	Gender
Age	Age at Booking
Residence Zip Code	Offender Residence as of run date
CaseID	Court Case Number
Arrest Date	Date of Arrest for the charge
Booking Date	Booking Date
Booking Agency	Arresting Agency
Charge	Charge Code
Charge Description	Charge Description
Level	Charge Severity
Sentence Date	Data the charge was sentenced
Booking Authority	Booking Type

Bail Amount	Bail Amount for the charge
Charge status	Status of the charge as of run date
Release Date	Release date from the booking
Release Reason	Reason or type of release
Race	Race/Ethnicity
Charge section and statute	Charge Section and statute (PC 459)
Charge Hierarchy	Charge Hierarchy Number
Offense group	Charge Category (Person, Property)
Offense type	Charge Grouping (Robbery, weapons, etc)
Brief Jail Mental Health Screen Result	MH need associated with the initial screening

Example dataset for Jail Booking and Release (link)

Example dataset for Competency/Incompetent to Stand Trial (link)

Table of statutes groupings and relative severity (link)

PROBATION

Variable Name	Description
Client ID	Unique Client ID
Case ID	Unique Case ID
Statute of leading case	Section and Statute
Supervision Start Date	Probation Start Date
Supervision End Date	Probation End Date
Caseload Type	Probation Caseload
Assessment date	Date of Assessment
Assessment type	Type of Assessment
Static Risk Level	General Tool
Dynamic Risk- Priority 1	Needs Assessment
Dynamic Risk- Priority 2	Needs Assessment
Dynamic Risk- Priority 3	Needs Assessment
Violation Petition Date	Date of Violation Petition
Violation Petition Type	Petition Type or Reason (New Crime, Technical Violation)
Violation Petition Disposition	Probation Petition Outcome (Revoked, Reinstated)
Conviction date	Date of new conviction

Example probation case and violation data file (link)

Example Probation Risk and Needs Assessment data file(link)

Example Pretrial Assessment (link)

CASE PROCESSING AND COURT DATA

Variable Name	
Case ID	
Case Filing Date	
Arraignment Date	
Arraignment Decision	
Case Disposition Date	
Case Disposition	
Hearing date	
Hearing Type	

Example Case Processing data file (link)

XYZ County has made it a priority to improve outcomes for those with behavioral health issues who interact with the justice system. Starting with a Stepping Up resolution in 2015, currently convening in the Data Driven Recovery Project (DDRP), the county has committed to better informing its strategies around this shared population. This plan is intended to lay out a framework for supporting the technical as well as use issues for XYZ County. This initial merging of data would use records from multiple system only for the purpose of aggregate analysis, no information sharing, or exchange would occur at the person level and any analysis would work to avoid the risk of re-identification using best practices and standards.

The strategy herein covers 3 areas:

- Approach for the creation of a county identifier
- Creation of a data warehouse for justice and behavioral health clients
- Outline of a data governance plan

Due to the complexity in sharing rules as well as analytic considerations, projects like this are rarely sustained, or started, with the intent of building capacity and plans for the long term. Through the DDRP, XYZ County has obtained the external resources as well as internal support to begin to develop a long-term technical plan which considers privacy issues as well as usage interests.

The general steps and timeline are as follows:

1. Engage data owners with behavioral health services, Sheriff's Custody/Jail facilities, and probation's community corrections on data extraction and feasibility.
2. Create and approve a county plan approved by department heads and county leadership to authorize the merging and analysis of shared data.
3. Develop data warehouse with shared records
4. Convene and engage a data governance workgroup to ensure each data owner's data is used appropriately, and processes are put in place for review
5. Engage county leadership on continuity planning and resource development to sustain data governance in anticipation of the end of the DDRP grant.

This document maps out a data merging plan for the purpose of data analysis and research, that will assist the county in answering key questions about those with behavioral health issues like:

- Prevalence in various parts of the justice system
- Jail recurrence and recidivism
- Types of services and events in both justice and behavioral health

- Clarifying definitions of levels and diagnosis associated with behavioral health issues

There is currently no common identifier across behavioral health and justice agencies making statistical analysis unreliable regard to many of the shared population. As the county begins to look at policy and practice options for clients across agencies, the need to merge select data fields is a fundamental first step to create baselines and develop a longer-term research and analysis strategy. Since this data is being used retrospectively, there is no data being passed between entities for service provision, and Personally Identifiable Information (PII) is only needed for the initial matching of records and will be deleted or de-linked early in the data processing.

The goal is for the data request to be intentional about what is shared by each data owner, merging the minimum number of fields from each data owner to reduce query and merging complexity, but still provide value in answering questions of interest.

The first section outlines an approach that masks client identifiers before any analysis has occurred, with the output being aggregated charts and tables based on the events in various data-owners' systems. The second covers the for the database merging and loading for analysis. The third section lays out how the county will limit the risk of re-identification and client security protocols will be developed through a process of data governance, that will create rules for both how data is used, as well as how its organized and presented. Standard procedures and norms will be created to ensure client data is kept secure, each data owner's aggregate data is presented accurately, and that aggregate reporting will not risk re-identification.

Its envisioned the initial analysis would cover data attached to adults with episodes and relevant justice events from 2016-2019 (4 years) across different systems ensuring a window into current practice, but also enough historical background to understand longer term trends.

The data owners in Phase 1 of the information sharing endeavor are:
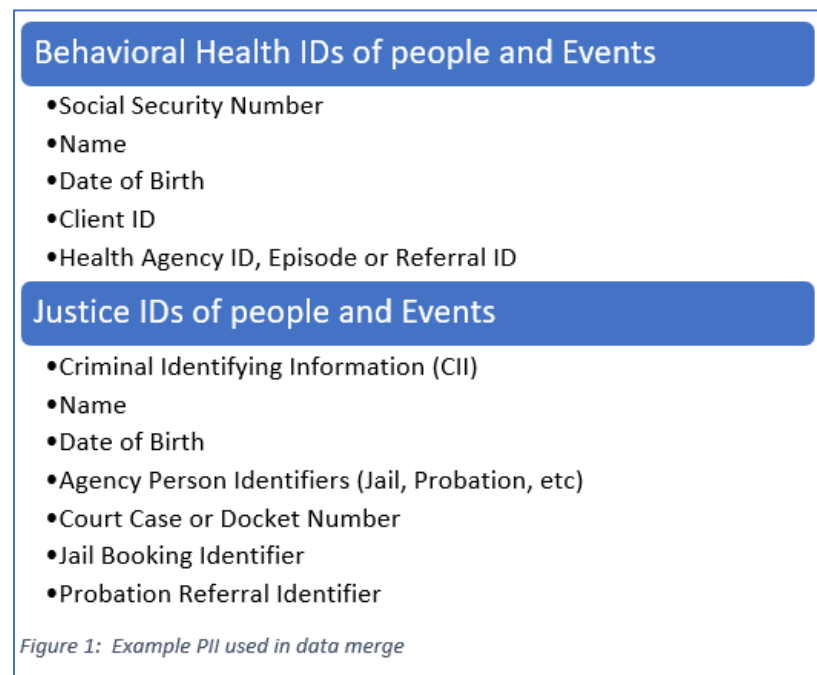
- XYZ County Behavioral Health
- XYZ County Sheriff's Department
- XYZ County Probation Department
- XYZ District Attorney

CREATION OF A COUNTY IDENTIFIER, ACROSS SYSTEMS

Both CORI and HIPPA offer exceptions for the use of PII for research and internal planning. Further, the use of the data in the current proposal is internal to the county, not a release of

information to a third party. This initial project plan lays out a strategy for merging data across agencies where California Penal Code § 13202[6] grants analytic/research provisions as does HIPAA under 45 CFR 164.501 and 42 CFR § 2.52[7]. However, both require the removal of PII. The PII envisioned for merging is listed in Figure 1, with priority given to common numerical identifiers and then direct identifiers like name and date of birth.

Behavioral Health IDs of people and Events
- Social Security Number
- Name
- Date of Birth
- Client ID
- Health Agency ID, Episode or Referral ID

Justice IDs of people and Events
- Criminal Identifying Information (CII)
- Name
- Date of Birth
- Agency Person Identifiers (Jail, Probation, etc)
- Court Case or Docket Number
- Jail Booking Identifier
- Probation Referral Identifier

Figure 1: Example PII used in data merge

Using an intermediary step that merges PII to build a master list of people in behavioral health and justice systems can then avoid explicitly sharing PII as merging and transformation would happen in a protected environment, to only use de-identified data for analysis. This merging of PII would occur using the SOUNDEX function, used to match names based on phonetic spelling, paired with the date of birth.[8]

---

[6] Notwithstanding subdivision (g) of Section 11105 and subdivision (a) of Section 13305 , every public agency or bona fide research body immediately concerned with the prevention or control of crime, the quality of criminal justice, or the custody or correction of offenders may be provided with such criminal offender record information as is required for the performance of its duties, provided that any material identifying individuals is not transferred, revealed, or used for other than research or statistical activities and reports or publications derived therefrom do not identify specific individuals, and provided that such agency or body pays the cost of the processing of such data as determined by the Attorney General.

[7] The HIPAA Privacy Rule establishes the conditions under which protected health information may be used or disclosed by covered entities for research purposes. Research is defined in the Privacy Rule as, "a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge." See 45 CFR 164.501 and 42 CFR § 2.52. A covered entity may use or disclose for research purposes health information which has been de-identified (in accordance with 45 CFR 164.502(d), and 164.514(a)-(c) of the Rule) and 42 CFR § 2.52 (b)(3).

[8] https://www.archives.gov/research/census/soundex.html

The merge of behavioral health and justice PII would then be "pseudonymized[9]" to make it identifiable only to a certain subset of database administrators on the "backend" of the system or deleted entirely. Merged data would not be passed back to the original data owner so no new data or identifiers would be added to the original data owners' submissions or data flow. The transformed data would be loaded into a data warehouse containing identifiers as pseudonyms, as well as selected fields from each data owner. The merged dataset would also anonymize any record locator or case file ID. Pseudonymization does not remove all identifying information from the dataset but merely reduces the clear relationship of a dataset with the original identity of an individual. The produced data warehouse resources would then be managed by workgroups/teams formed through the data governance process, working across agencies. Figure 2 shows how the initial loading of PII creates a universe of people common to all data owners, which is then merged with event and episode identifiers, after which PII is given a pseudonym, such that the final dataset contains no PII.

---

[9] To pseudonymize a data set, the additional information must be kept separately and subject to technical and organizational measures to ensure non-attribution to an identified or identifiable person.
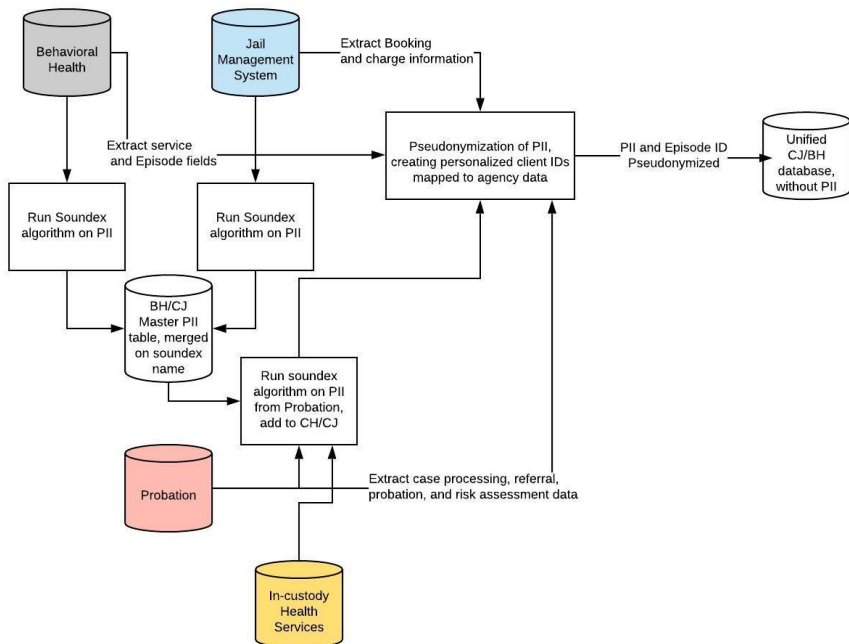
Figure 1: Flow Chart of Data exchange, transformation, and loading of Data Warehouse

CREATION OF BEHAVIORAL HEALTH AND JUSTICE AGENCY DATA WAREHOUSE

Once a common identifier is established, with PII removed as noted above, a limited set of fields would be extracted from the following databases to create a single data warehouse for analysis that could look across agencies at shared clients, but not pass any PII. Since each data owner has unique workflows and data structures, the initial data extraction would attempt to create the basic flow of people through each system over time, then carefully build out common baselines and analysis. A number of these fields are considered "indirect identifiers", so would need to be used with caution when time to report or aggregate to mitigate the risk of "re-identification". It is assumed the jail's demographics facts would be used, mainly because the use of live scan/fingerprints may be better than self-reported information elsewhere. Appendix A notes the fields of interest.

The databases of interest initially are:

- Behavioral Health Client Management Systems (Community) (EHR name)
- Behavioral Health Client Management Systems (In-Custody) (EHR name)
- Jail Management System (System name)
- Probation Systems (Case management system name and Risk/Needs Tool)

The data warehouse and analysis would initially be overseen by XYZ for the term of July 2021, then other resource in the county would need to be identified. The contractor would work with

the data governance group envisioned below to develop, problem solve, and analyze the data available. This ensures capacity building with county staff, as well as a trusted third party in the development of norms for the use and handling of the merged data.

The data merging will happen in two phases:

- Phase 1: Developing and understand each data owner's data and process, with no PII present. The only identifying information would be database system identifiers. This would allow analytic work to begin, while discussions on privacy or sharing issues are resolved. Its envisioned that specific data extractions would be created and reside separately across the three data owners. Data would be prepared as well documented during this phase.
- Phase 2: Request personally identifying information from each data owner for merging across systems. This merge would be done using computer code, so no manual record observation is required. There will likely be a need for quality control, but this would be done by the external analyst, not a specific data owner to maintain anonymity.

By running this project on parallel tracks with the approval of the sharing plan, as well as the development of individual data requirements, the hope is to reduce the time waiting. When merging is approved the county would have the technical database resources in place to move to Phase 2.

DEVELOPMENT OF THE DATA GOVERNANCE PLAN

Data governance is the guiding principle that manages the data assets and uses of data functions. Data is an evolving set of functions for integrating behavioral health and justice data, and therefore any aggregation requires expertise and vision on where to take the data, process for setting priorities on adding or defining data elements, advising on the uses of collected data, and deciding on quality control methods across agencies. The gathering and management of behavioral health and justice data into a single warehouse requires an improvement in infrastructure to guarantee availability, usability, integrity, and messaging. A data governance program includes a governing group, a defined set of procedures and activities, and a plan to execute the procedures. This means putting personnel and organizational structure in place to achieve this goal. The varying standards of health and justice mean it's more important than ever to document and clearly define the needs and uses of agency data. Any analysis would work to avoid the risk of re-identification using best practices and standards.

Data governance is a set of processes that ensures that important data assets are formally managed. Data governance also describes an evolutionary process, altering the county's way of thinking and setting up the processes to handle information so that it may be utilized by the entire county.

Data governance can be planned, managed, and implemented through a two-level structure ensuring a county defined mix of executive level support and sponsorship as well as subject matter experts.

- Existing executive level support could be maintained through the existing Stepping Up framework, or something more permanent
- A Research and Development Workgroup could provide strategic direction and ensures the data governance efforts address all relevant analytic demands. It manages data governance as an integrated program rather than a set of unconnected projects. Its strategic goals are to prioritize analysis efforts, communicate with or represent county data owners, and direct long term improvements in collection and integration.

A data roadmap covers the implementation of guidance laid out from an executive or strategic group. This group can cover both research/development goals around using the data for analysis and applied topics but can also have operational or workflow expertise to ensure approved users get access to the correct data.

A Research and Development Workgroup can be tasked with managing the uses of the resulting merged and de-identified dataset. Once this datawarehouse is developed, it can be used to answer important and prevalent questions about the County's population of those with behavioral health issues and justice involvement. Further, it can be used for questions that relate to a single agency as well since the purpose of the data warehouse is to get easy access often complex datasets. The Workgroup can develop a process for receiving data requests, providing the appropriate data to answer those requests, and managing relationships with the requestors. County Behavioral Health and Justice agencies receive numerous questions and requests for information from sources internal and external to the County government regarding this population and the dataset created through the DDRP can be an important asset for answering these questions efficiently, ultimately informing research studies, evaluations, and efforts to improve program provision. These uses can be for analysis within the county or for using external partners.

Three common use cases for a process to share this data are:

1. One-time data request, such as
   - A community-based organization interested in identifying the population it serves and learning more about its characteristics to develop effective programs to address the population's needs
   - A staff member of a County agency participating in the DDRP who wants to know more about the population that accesses the behavioral health and justice systems to inform how best to serve this population
   - A County Supervisor interested in trends regarding the behavioral health/justice-involved population to inform budget allocation decisions
2. An external research partner, such as
   - A representative of an established research organization or university interested in studying the County's behavioral health/justice-involved population and the effectiveness of various interventions the County employs
3. A contracted evaluator such as:
   - An evaluator or researcher under contract with a County agency to evaluate a County program or intervention
   - A researcher under contract with a County agency to validate an assessment used by County behavioral health and/or justice agencies

- A researcher serving as an external evaluator for a grant, who is either recruited later, or part of the initial grant writing process

The elements the Workgroup should consider in establishing a data governance process to manage data requests include the following:

FAMILIARITY WITH THE DATASET

The Workgroup should be familiar with the data contained in the dataset – the fields that are included, any limitations of the data, and any risks of potential re-identification that may exist in the dataset. This knowledge will inform how the Workgroup responds to data requests, for example, determining whether to provide the data as requested, provide some of the data requested to protect confidentiality, or provide some or all of the data requested with an explanation of what the limitations are of the requested data for answering the requestor's question(s).

REVIEW OF FREQUENTLY REQUESTED DATA

Before establishing a data request process, the Workgroup may want to identify the most requested data related to the behavioral health/justice-involved population. This can be achieved through a review of the data requests that the participating agencies have received over the previous 2 years, depending on the volume of requests received annually, and should include enough data requests to identify trends. The result of this review will be a summary of the data elements most frequently requested, the most common type of requestor (internal county agency, external researcher, student, community based organization, etc.), and the most common ways the requests are received (through an agency website, a public information request, directly to a staff member, etc.). To see an example of the form to help pull these together as well as the output of this, its linked here.

CREATION OF AGGREGATE DATA THAT CAN BE SHARED PUBLICLY

With the results of the review from Step 2 above, the Workgroup can identify aggregate data tables and charts that can be prepared from the dataset and can answer the more frequently received data requests and questions. These tables and charts can be shared publicly on a County website, with links on each agency's site directing visitors to this information. This will likely reduce the number of requests each individual agency receives related to this population and will increase the County's transparency and collaboration with data-driven research and community organizations. An Example of available tables are here.

CREATION OF A DATA REQUEST PROCESS

While data requestors should first be directed to the publicly available data tables and charts created in Step 3, they should also be offered a way to request data that is not included there. This may include more nuanced information than is provided in the public data and/or individual-level data or data aggregated on different metrics. The Workgroup can create a process for submitting these requests, such as an online form, and should identify an agency or individual from the Workgroup to receive these requests. This process should include how the requests are reviewed and how it is decided whether to comply with the request or not, depending on factors such as: resources (i.e., staff time) needed and available to comply with the request, risk of re-identifiability in requested data, purpose of research / data request and its alignment with County strategic goals, etc. Requests may be reviewed periodically at Workgroup meetings or on an as-needed basis in in-person meetings or through another process decided upon by the Workgroup. An example of the workgroup intake form is here for internal requests by county agencies, and here for external partnerships.

TRACKING AND MANAGING DATA REQUESTS

Once a process is established for determining which data requests to fulfill, the Workgroup must establish a process for fulfilling the request, including: who will compile the requested data; what type of review, if any, will be required before sharing the data; what type of agreement (i.e., Memorandum of Understanding or Data Use Agreement) will be required in order to share the data, how the data will be shared (i.e., on a secure site), who will be the point of contact between the Workgroup and the requestor, and whether the requestor's resulting report or study will be required to be shared with the Workgroup. These items may vary based upon the type of data requested, with aggregate data requests likely requiring less stringent data sharing procedures than individual-level data requests. A standard data sharing agreement is here.

A system should be established to manage and track the data requests, the Workgroup's responses, the agreements executed, the requestors' contact information, and the reports or studies that result from the data sharing. An example of this form is here.

Creating a data sharing protocol and process as described above will ensure that the dataset created through the DDRP is accessible and available to be used internally and externally to enhance, evaluate, study, and improve services and outcomes for this population. In addition, by identifying the most commonly-requested data and providing publicly available data to address those requests, the number of ad hoc data requests the participating agencies receive regarding this population and the staff time it takes to comply with them will be reduced. Further, it will enhance consistency of response since the data would be pulled in the same way through a clearly defined, and automated process.