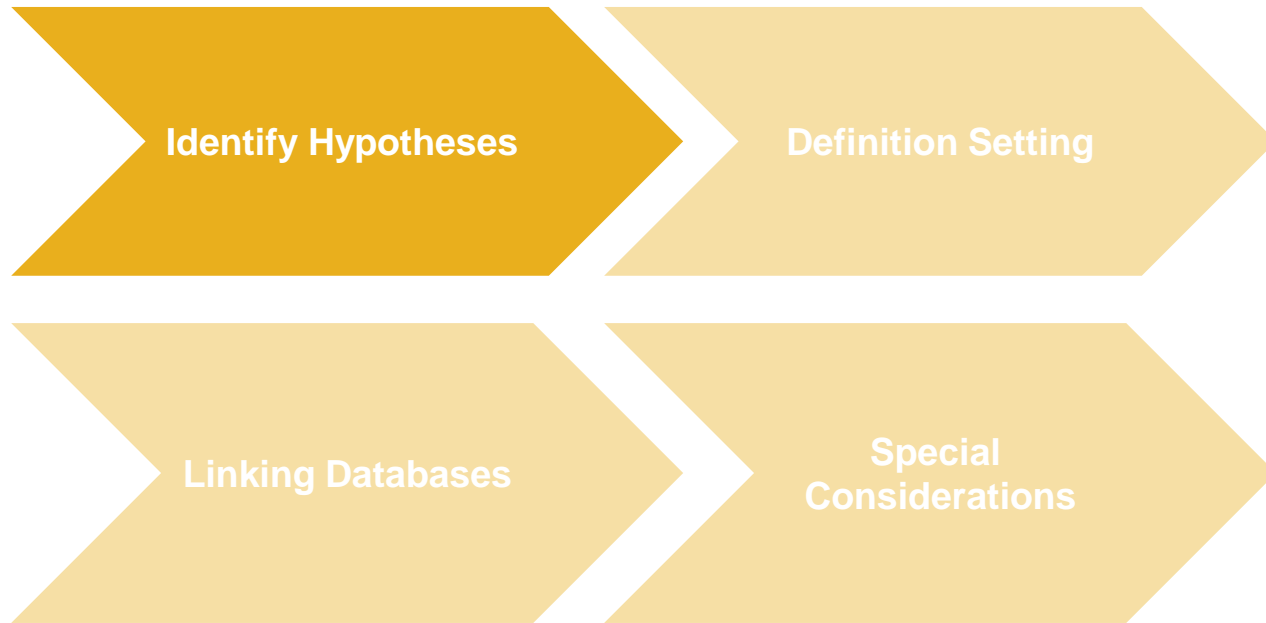


DEEP END DATA CHALLENGES

**ANTICIPATING AND OVERCOMING CHALLENGES IN DATASET CREATION AND
ANALYSIS**

THE ANNIE E. CASEY FOUNDATION, MARCH 12-13, 2015

Creating Your Dispositional Dataset



Identifying hypotheses is a critical step in guiding data collection.

Guiding hypotheses help clarify the most important data to the project



Identify hypotheses about how youth arrive at out-of-home placement

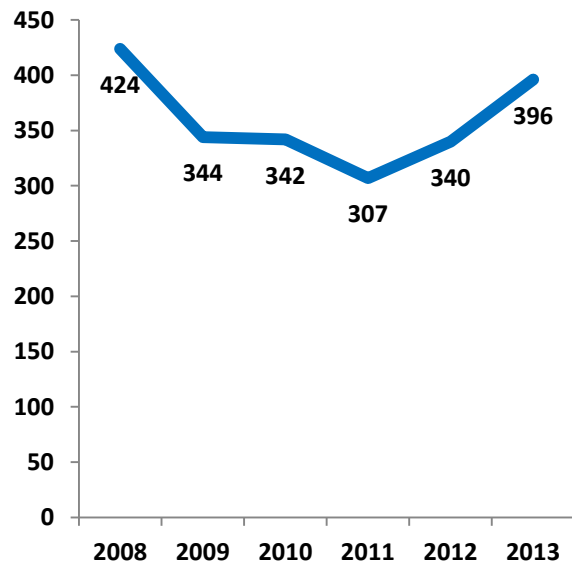
Such as...

- Youth with ***more serious offenses*** are more likely to be sent to placement
- Youth ***classified as higher risk*** are more likely to be sent to placement
- Youth with ***identified mental health needs*** are more likely to be sent to placement
- ***Older youth*** are more likely to be sent to placement
- Youth sent to placement based on ***probation violations*** have more serious underlying charges than other youth placed on probation

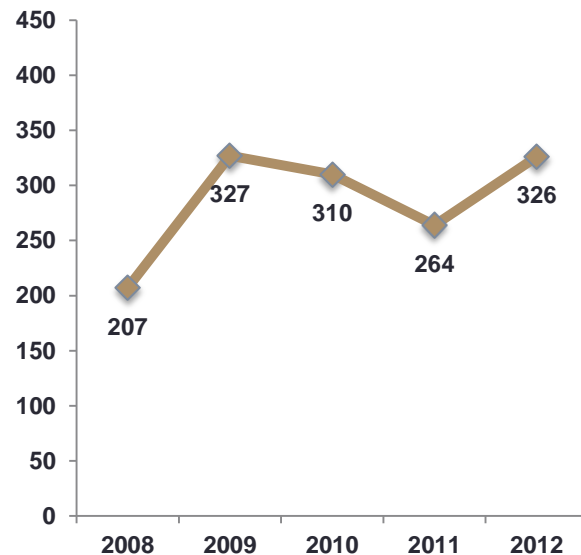
An example from one of the initial Deep End sites shows that no number speaks for itself: context and points of comparison are everything.

- Initially, the site's analysis resulted in a large set of charts like these, that each look at one thing. This was an important step in that it provided the grist for all subsequent analysis. But it was hard to interpret what was happening based on stand-alone numbers.
- What do you think was happening in this community with respect to the handling of VOPs?
- What would your next step be to make this information more comprehensible?

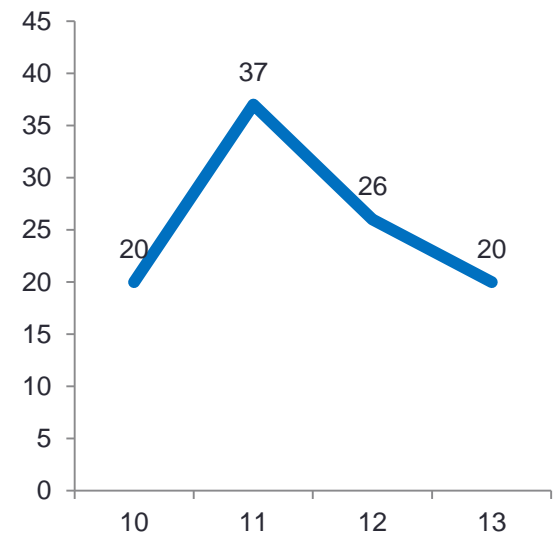
VOPs Filed by Fiscal Year



VOPs Disposed by Fiscal Year

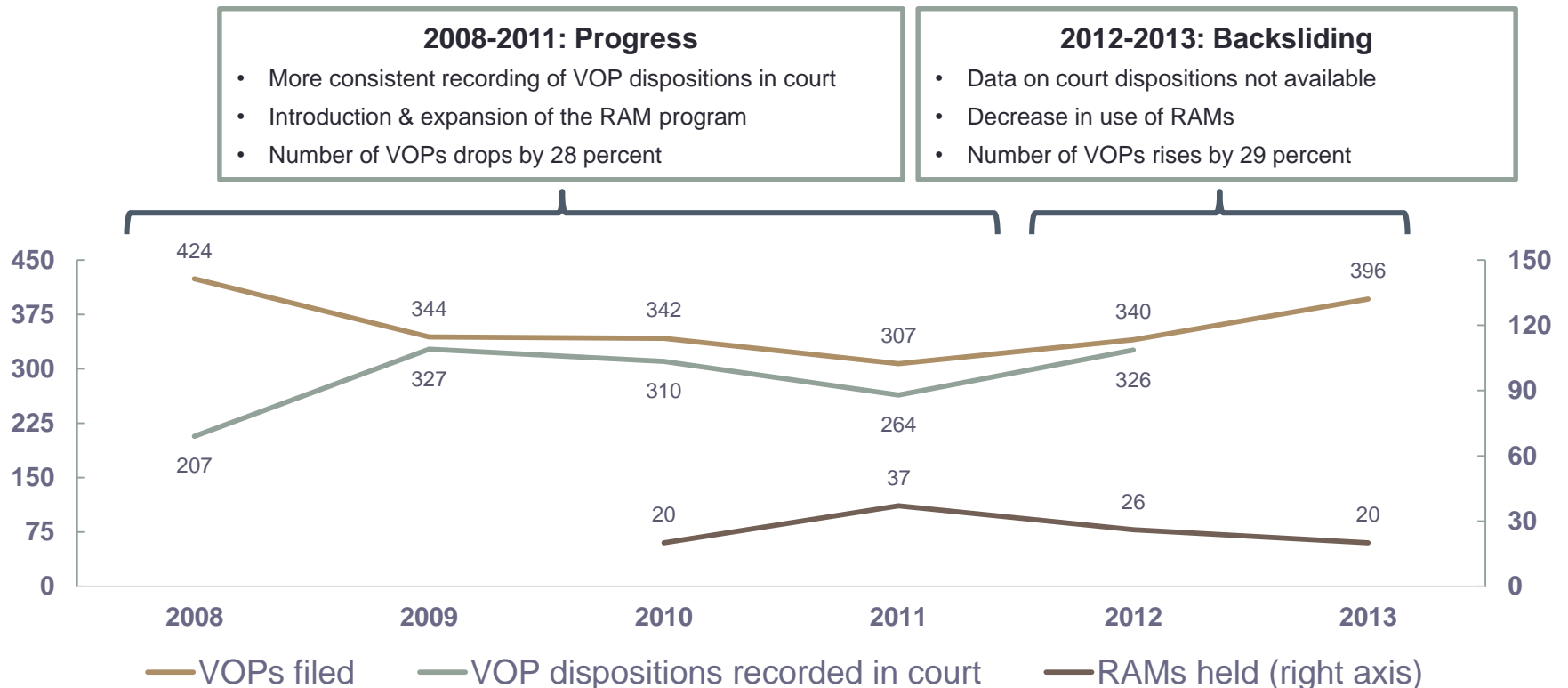


Revocation Alternative Meetings by Fiscal Year



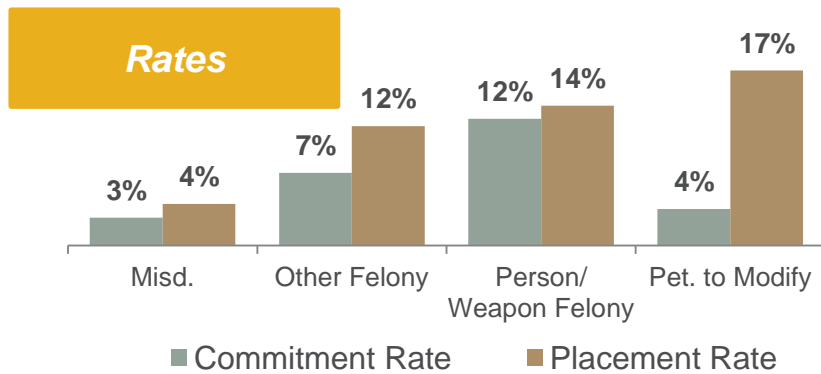
Getting this information onto the same timescale, and comparing trends over time, illuminates a potentially critical issue.

- It appears that concentrated efforts and reforms helped to drive down the number of VOPs from 2008 to 2011 – but that progress has started to reverse since then.
- Determining what accounts for this reversal, and how the jurisdiction could get back on track, could become an important strategic driver for the site’s Deep End work.

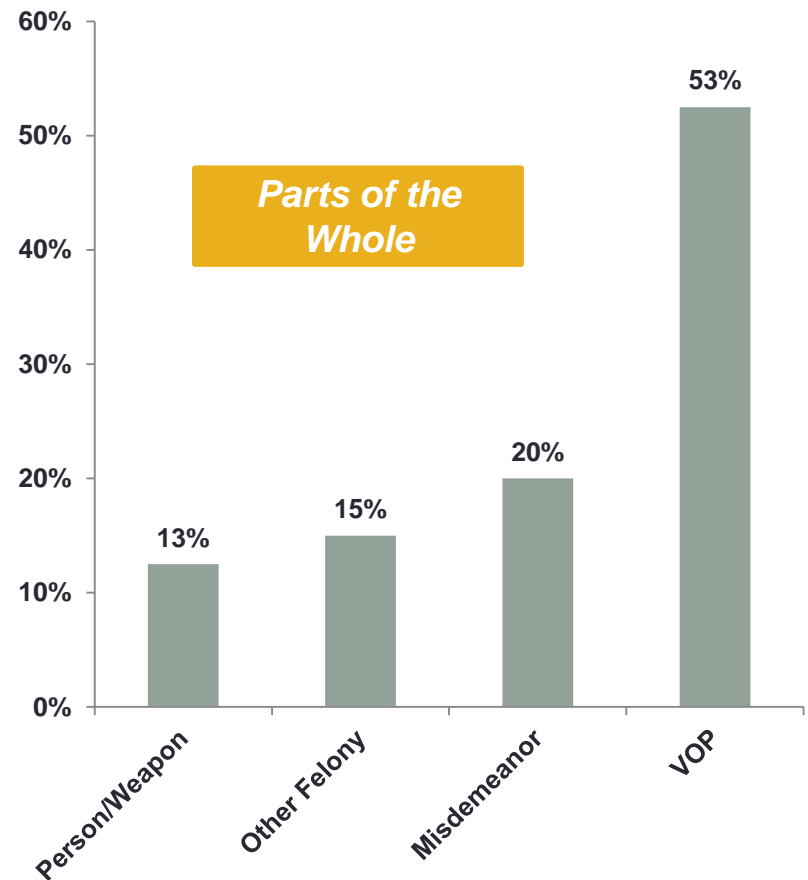


Graphically, the first cohort of Deep End sites used a variety of displays to show what they learned about their hypotheses.

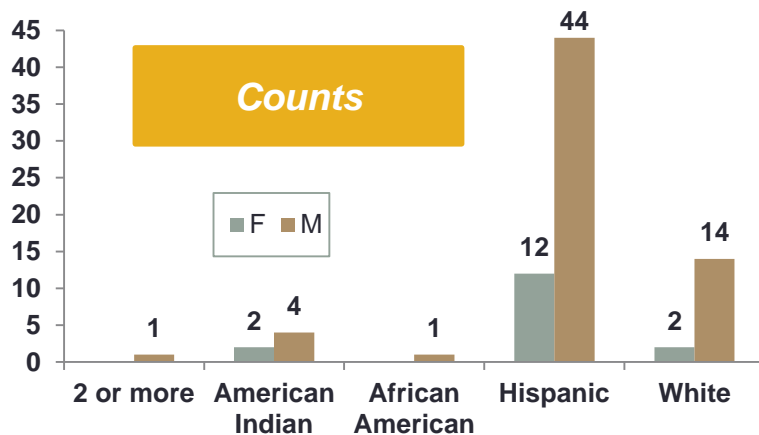
Rates of Commitment & Placement by Offense



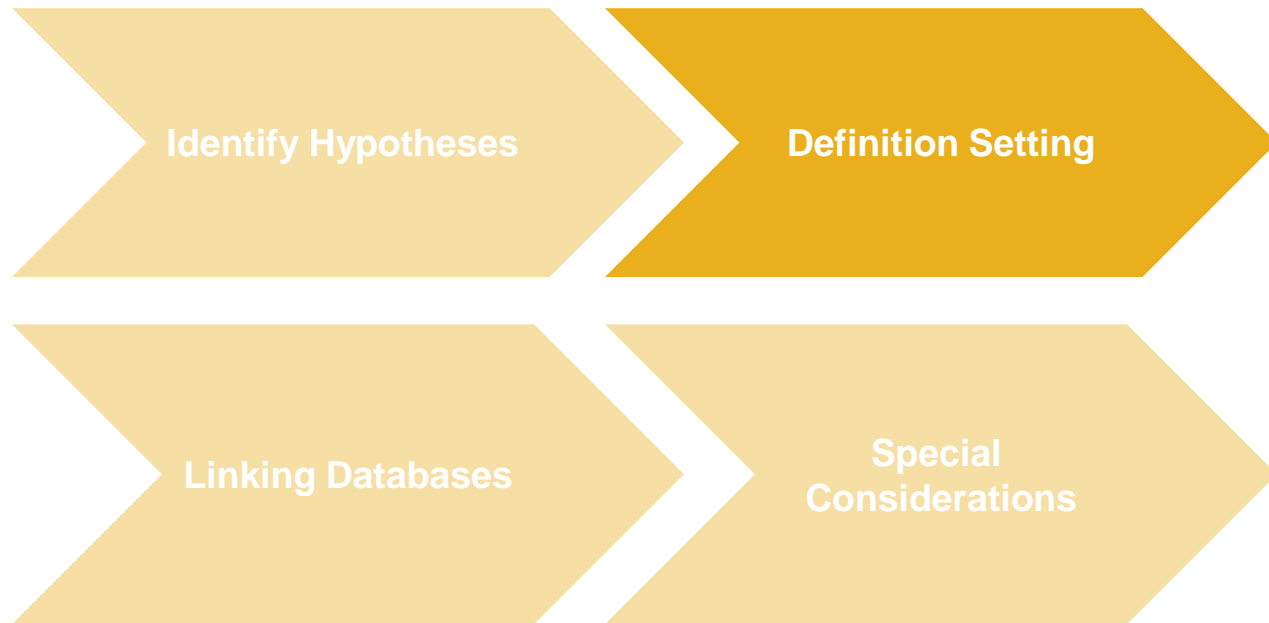
Commitments by Most Current Offense



Commitments by Race & Gender



Creating Your Dispositional Dataset



Defining the Target of the Analysis: Placements

Different jurisdictions, different agencies within a jurisdiction, and even different people within an agency may define “placement” in different ways

Deep End Analysis should include all youth who are:

1. Placed outside the home
2. As the result of a delinquency
3. Pursuant to a court order that prohibits the youth from leaving

* Regardless of local or state custody, public or private operation, security-level, or name used to describe the facility

Defining the Target of the Analysis: Disposition Decisions

What is a Disposition?	How does database construction impact your analysis of dispositions?
Resolution by the JJ system of a legal matter involving one or more alleged delinquent acts, status offenses, or violations of probation by a young person under the system's jurisdiction	Ideally, each disposition will be linked to one or more specific charges.
DO include "intake dispositions" - e.g., diversion.	CHALLENGE: What if dispositions are only linked to complaints?
DO NOT include procedural court actions that may be coded as dispositions, but do not actually resolve the case – e.g., "continuance."	CHALLENGE: What if dispositions are linked to multiple complaints?
DO dig deeper into critical definitional issues – e.g. does "commitment" always mean placement, and does "probation" always mean supervision in the community?	CHALLENGE: Depending on the answers, you may need to get creative about filling the gaps

Defining the Target of the Analysis: Disposition Decisions

What is a “disposition decision”?

1. Decision is about one youth at one point in time
2. Decision is based on the disposition of one or more delinquency, VOP, or status charges
3. Decision results in the conclusion of formal proceedings or the initiation of a juvenile justice intervention for that youth



Examples of results that should be counted as “disposition decisions”

- | | |
|---|---|
| <ul style="list-style-type: none">• Transfer/waiver to adult court• Commitment to state custody• Other out-of-home placement• Probation or other community supervision• Provision of court-ordered services | <ul style="list-style-type: none">• Imposition of fines, court-ordered community service, or other alternative resolutions• “Informal probation”• Diversion from formal processing• Dismissal, withdrawal of charges, exoneration, or other court action ending the youth’s legal jeopardy |
|---|---|

Key Attributes of Disposition Decisions

Agreeing on common definitions will ease data collection, analysis, and interpretation.

Variable	Questions to ask about how your jurisdiction captures this data
Race	<ul style="list-style-type: none">• Are race and ethnicity coded separately?• Do youth self-report or does someone else make the determination?
Disposition type and groupings	<ul style="list-style-type: none">• What disposition options are available in your jurisdiction?<ul style="list-style-type: none">○ E.g., diversion, probation, commitment, other out-of-home placement, etc.• Can you aggregate them into general categories?
Offense Type	<ul style="list-style-type: none">• How are offenses categorized?<ul style="list-style-type: none">○ E.g., felony/misdemeanor, violent/non-violent, person/property/drug, etc.?
Most serious charge	<ul style="list-style-type: none">• How is the most serious charge determined/defined?<ul style="list-style-type: none">○ Is there a severity table?<ul style="list-style-type: none">○ Does the severity table make sense to stakeholders?○ Do related systems use the same severity ranking?
Offense	<ul style="list-style-type: none">• From which decision point should you draw offense information?<ul style="list-style-type: none">○ Analyses of arrest, referral, diversion, and detention → Alleged offense○ Analyses of formal filings → Petition charge○ Analyses of dispositions → Adjudicated charge

Evaluating Data Quality & Availability

It can be very helpful to systematically assess your data sources, and keep track of your findings.

Potential Tool: Data Diagnostic Table for Dispositional Dataset

Type of Data Needed	Current Availability Status			Data Source	Perceived Data Quality
	Readily and Immediately Available	Available, but Messy	Not Available		
Demographics: Race	Yes			Social Services database	Missing 5% of data
Violations of Probation		Yes		Probation database	Missing 35% of data
Risk Assessment			Yes	Probation case file	Not available electronically

Questions to help determine data quality include...

- How frequently are data entered?
- Are data coded consistently?
- What is the percentage of missing data?

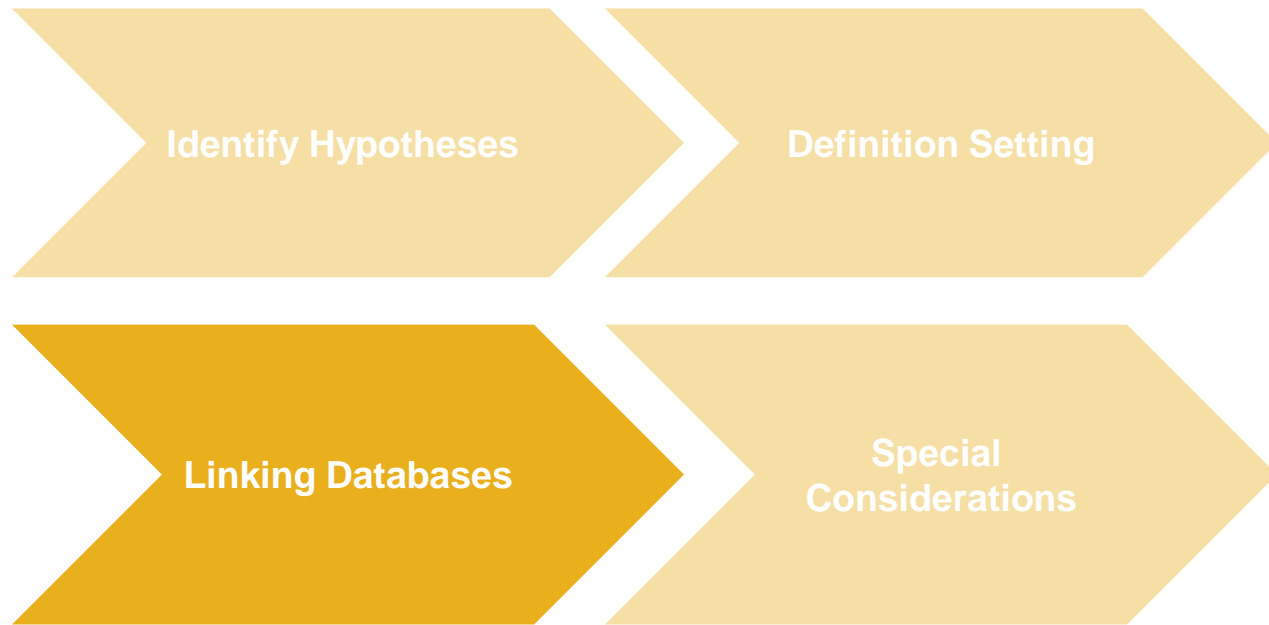
Securing the Data You Need

Don't be the person your data providers hate to see coming:
be as precise as you can with your data requests.

Question	Example
How to define each data point?	Race: White, Black, Asian, Other Charge Severity: Felony, Violent felony, Person-felony, person-based misdemeanor
Do you know how to interpret codes?	M/F is easy, but what does RDDDL mean? TA10?
How to be clear about the time period?	All dispositions in 2014 v. All dispositions arising from referrals in 2014
What's the unit of analysis?	One row per charge v. one row per case v. one row per youth
What identifiers should be included to protect confidentiality, allow accurate matching datasets, and preserve your ability to investigate individual cases?	Youth ID, case ID, complaint ID, etc.

(Some stereotypes are true: it never hurts to show up with cookies & donuts, either)

Creating Your Dispositional Dataset



Data about the Deep End comes from multiple sources.
You will need to link records across those data sets.

Identifiers are essential for linking datasets from multiple sources

- For placement analysis, the ideal identifier is an individual-level identifier (e.g. youth identifier)
- This may be challenging for some jurisdictions – as an alternative, a case-level identifier (e.g. case number, referral ID) is still helpful
- Datasets may contain multiple identifiers at different levels, e.g. individual and case-level
- In the worst case, you may need to do probabilistic (“fuzzy”) matching
 - This is painstaking & laborious, but worth the effort if there’s no other way
 - But if you need to go this route, take it as a sign that you really need to create a more consistent identifier

**Data matching is rarely easy or error-free –
be prepared to tolerate some messiness**

The reliable identifiers you have available will influence what kinds of analysis you will be able to conduct.

Unit of analysis

Individual-level

Case-level



Example Identifiers

Unique youth ID, combination of full name and date of birth

Docket number, case number, arrest number



Attributes available to be examined

Age, race, gender

Offense types, charge severity

Especially when you are drawing on data from multiple sources, it is often necessary to aggregate, or “roll up”, some pieces of information.

Aggregating data involves consolidating several pieces of information into a single data element – usually based on a common person or a case identifier – to create more useful measures than might be available in the administrative data

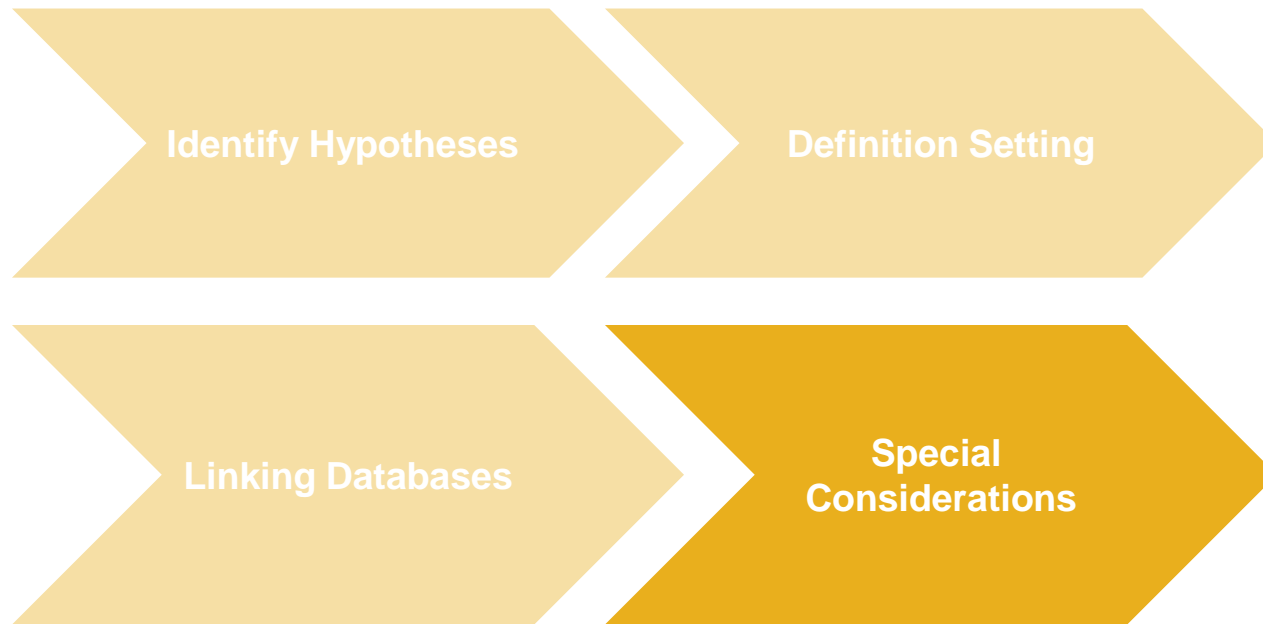
Youth ID	Arrest Date
12345	1/2/2010
12345	6/4/2010
12345	8/6/2010
12345	3/8/2011
12345	4/10/2011



Youth ID	Number of Prior Arrests
12345	5

Worth noting: this is much the same procedure that many jurisdictions need to apply to come up with their set of “disposition decisions” – i.e. they need to roll up the dispositions for multiple offenses or petitions into one unit of analysis, typically based on the youth ID and the disposition date(s)

Creating Your Dispositional Dataset



Probation Recommendations

The P.O.'s recommendation is an important, but often overlooked, influence on dispositional decision-making.

Probation departments often make recommendations to the court as to what disposition would be appropriate.

- ▶ If your jurisdiction collects data about dispositional recommendations, you should include it in your analysis
- ▶ If not, you should think about starting to collect it

Obtaining and interpreting data on probation recommendations can be difficult for a host of reasons:

Court practices	Courts that make heavy use of plea bargains and other opaque decision making methods	Lack of standards or guidance to front-line probation officers that could help them to make solid recommendations
Data collection practices	High caseloads & time-consuming assessment procedures, that make it difficult to provide timely, well-informed recommendations	Lack of standards or guidance to front-line probation officers that could help them to make solid recommendations

Violations of Probation

In most jurisdictions, VOPs represent one of the largest doorways to out-of-home placement.

What Leads to a VOP

An arrest on a new offense, which was processed as a VOP rather than a new charge

A technical violation – failure to comply with conditions or probation

To dig deeper into VOPs, you want to build your dataset in a way that allows you to...

Link the VOP back to the most serious offense underlying probation

Determine how long the youth was on probation prior to the violation

Categorize the violation itself, e.g. new arrest vs. technical (type of technical)

Risk Assessments

Pre-dispositional risk assessments can heavily influence dispositional decision-making – yet they are often recorded completely separately.

There are important differences in how risk assessment is handled at detention and at disposition – but from a data standpoint, much of what you’ve learned from JDAI is still applicable.

When Risk Assessments Are Used At...

Detention - measures short-term risk of flight and re-arrest

Disposition – measures youth’s long-term risk of re-entering the system

Common Challenges

Youth may have multiple assessments

The jurisdiction may not collect this information - think about how you might do this in the future

The information may exist, but it is not connected to other important information - refer back to linking datasets section

Placement Moves & Length of Stay

The focus of local Deep End work has been on the major pathways into placement (disposition and probation). But in some jurisdictions, placement moves & LOS matter even more than admissions.

	Why it matters	Data challenges
Placement moves & changes in security levels	<ul style="list-style-type: none"> ▪ Movements between placements, especially when they involve changes in security levels, can be a surprisingly important driver of admissions into placement 	<ul style="list-style-type: none"> ▪ Changes in placement or security levels may not be tracked as closely as dispositions ▪ These may not tie back to the original case <ul style="list-style-type: none"> ➤ Where identifiers are lacking, dates can be used to string together some of the key relationships
Lengths of stay in placement	<ul style="list-style-type: none"> ▪ Local jurisdictions increasingly have control (or at least influence) over LOS, whether in local placements or even in state commitments ▪ Research shows how much harm excessively long stays can do to youth outcomes <ul style="list-style-type: none"> ➤ If we want better outcomes, we have to be thinking about it 	<ul style="list-style-type: none"> ▪ Placement providers (or state agencies in the case of committed youth) are not always forthcoming with this data <ul style="list-style-type: none"> ➤ Implications for oversight & performance management ▪ Complicated by placement moves & security level changes ▪ May require linking data, not only between providers and the JJ agency, but also between providers and other providers

Don't Let the Perfect Be the Enemy of the Good

Go into the assessment knowing that you will leave some questions unanswered and some work undone.

Think of the assessment in phases

- Expect that each phase will leave unanswered questions to be addressed in future phases
- Sometimes the most useful findings in an analysis are the questions that it raises – because a problem clearly defined is half-solved

Call out the data challenges you encounter for future improvement

- Data needed for analysis are not available
- Identifiers are missing from datasets, and even creative approaches aren't enough to create the needed linkages
- Low match rates across datasets

Reflect on the process to prepare for the next set of challenges

- Start thinking about how you'll investigate questions about recidivism and other outcomes; and how to gauge the effectiveness of different dispositional alternatives



THE ANNIE E. CASEY FOUNDATION

Developing solutions to build a brighter future for children, families and communities

www.aecf.org