

# Building a Sustainable and Replicable Approach To Estimating the Prevalence of Youth Homelessness: *A Community Guide to Linking Administrative Data*

NOVEMBER 2024

## CENTER FOR POLICY RESEARCH

Lanae Davis, MPA

Kayla Christiani, MS

Ali McLagan, MA

## COLORADO EVALUATION AND ACTION LAB

Elysia Versen, PhD

Whitney LeBoeuf, PhD

Christian Belcher, MPA

Erin Wickerham, MPH

# Acknowledgements

The authors would like to thank the partners, policymakers, practitioners, and youth with lived experience who supported this work. Their expertise, guidance, and insight were invaluable in conducting this study and creating this guide. They include Colorado Department of Human Services, Colorado Department of Education and the Colorado Department of Local Affairs, Division of Housing, Office of Homeless Initiatives, Denver Public Schools, along with Colorado’s Network to End Youth Homelessness, made up of Colorado’s Runaway and Homeless Youth Service Providers, and each of Colorado’s CoC’s.

The authors would like to thank Dr. Joshua Barocas and his team of researchers at the University of Colorado, School of Medicine, Division of General Internal Medicine, Aurora, CO for their contribution to the multi-system prevalence estimates.

## Suggested Citation

Davis, L., Versen, E., McLagan, A., LeBoeuf, W., Belcher, C., Wickerham, E., & Christiani, K. (2024). *Building a Sustainable and Replicable Approach to Estimating the Prevalence of Youth Homelessness: A Community Guide to Linking Administrative Data*. Center for Policy Research.

<https://centerforpolicyresearch.org/publications/building-a-sustainable-and-replicable-approach-to-estimating-the-prevalence-of-youth-homelessness-a-community-guide-to-linking-administrative-data/>

---

*The work that provided the basis for this publication was supported by funding under an award with the U.S. Department of Housing and Urban Development. The substance and findings of the work are dedicated to the public. The authors are solely responsible for the accuracy of the statements and interpretations contained in this publication. Such interpretations do not necessarily reflect the views of the Government.*

# Contents

<b>Purpose of the Community Guide</b> .....	<b>4</b>
Iterative Approach.....	5
Cross Cutting Approaches.....	6
<b>Defining Study Goals</b> .....	<b>7</b>
The Power of Linking Administrative Data .....	7
Type of Estimates.....	7
<b>Identifying Potential Administrative Data Sources</b> .....	<b>9</b>
Defining Youth Homelessness .....	9
Considering the Contribution of Each Data Source to Study Goals .....	11
Examples of Administrative Data Systems .....	12
<b>Prioritize Administrative Data Systems</b> .....	<b>14</b>
Cultivate Champions Within Each System .....	14
Identify Systems, Partners, and Primary Contacts at Partner Agencies .....	14
Trust-Building Strategies with Organizations and Partners .....	14
Practical Feasibility of Linking Data from Each System .....	16
<b>Prepare and Link Data</b> .....	<b>17</b>
<b>Conduct Analysis</b> .....	<b>18</b>
Analytic Decisions.....	18
Types of Estimates: Known and Unknown .....	19
Incorporating Youth with Lived Experience in Data Analysis .....	20
<b>Centering the Voices of Lived Experts Through an Equity Lens</b> .....	<b>22</b>
<b>Frame the Message and Communicate Findings</b> .....	<b>23</b>
<b>Conclusion</b> .....	<b>24</b>
<b>References</b> .....	<b>25</b>
<b>Appendix</b> .....	<b>26</b>
What is LINC? .....	26
Benefits of LINC Collaborative .....	27

# Purpose of the Community Guide

Without an accurate estimate of youth homelessness, systems cannot be responsive to the needs of youth who are at risk of or are currently homeless. Systematic and rigorous information on (1) where homeless youth are located, (2) the number of youths experiencing homelessness, and (3) the barriers to ending their housing instability equips communities with the power to prevent and end homelessness.

Accurate data can lead to more appropriate allocation of resources and prevention efforts (Troisi et al., 2015). This may be especially true in geographic areas that have been historically more challenging to generate estimates for, such as rural areas (Boullion et al., 2022).

The study highlighted in this guide was conducted by the Center for Policy Research and its partners at the Colorado Evaluation and Action Lab at the University of Denver and the University of Colorado, School of Medicine, to:

- Build a sustainable and replicable approach to more accurately estimate the number of youth ages 14-24 experiencing homelessness in Colorado;
- Better understand the attachment of youths experiencing homelessness to major support systems (i.e., education, homeless services, and child welfare); and
- Learn more about the characteristics of youths experiencing homelessness, including the support services they access.

This community guide details a framework developed by the research team to inform state and local agencies on how to generate sustainable and replicable estimates of youth homelessness. The guide provides information on how to approach linking administrative data across systems to more rigorously estimate the prevalence of youth who have experienced homelessness.

The purpose of the Community Guide is to provide state and local agencies with a resource for building a sustainable approach to data sharing, analysis, and reporting across organizations. This guide focuses on the target population of youth experiencing homelessness; however, the guide is intended for a broad audience including government agencies, community-based organizations, and researchers to use when the ability to link cross-system data is available. The guide can provide researchers, policy-makers and practitioners with a tool to use in supporting similar efforts with other target populations of interest including public health, and criminal justice. In this guide, there is a framework for how to define study goals, identify administrative data systems, prioritize administrative data systems, select identity resolution approaches to link data across systems, conduct analysis, frame and communicate findings, identify and engage with key partners who can serve as champions for generating rigorous estimates of youth homelessness, and include youth with lived experience in the process. These steps are intended to be an iterative process, so that study goals can be expanded or refined based on learnings at each step in the process.

Illustrative examples from the study are used throughout this guide to demonstrate how the recommended approaches have been actualized by using administrative data from state and local agencies to estimate the youth homelessness population first in Denver, Colorado and then expanded to a broader geographic area statewide. The study used data from three systems - education, child welfare, and homeless management information systems - to estimate the prevalence of youth homelessness in Colorado and interviews with

youth with lived experience of homelessness and professionals. The study included qualitative data from interviews and focus groups with youth and professionals. The full study report provides more details about the analysis and narrative approach and can be found [here](#).

## Iterative Approach

This section provides details about the approach this study took to estimate the prevalence of youth homelessness in Colorado. The iterative approach consists of the following seven steps with two cross-cutting approaches that are incorporated throughout the process. The iterative approach is outlined below and are explained throughout this section by presenting the questions used to frame each step in the process.

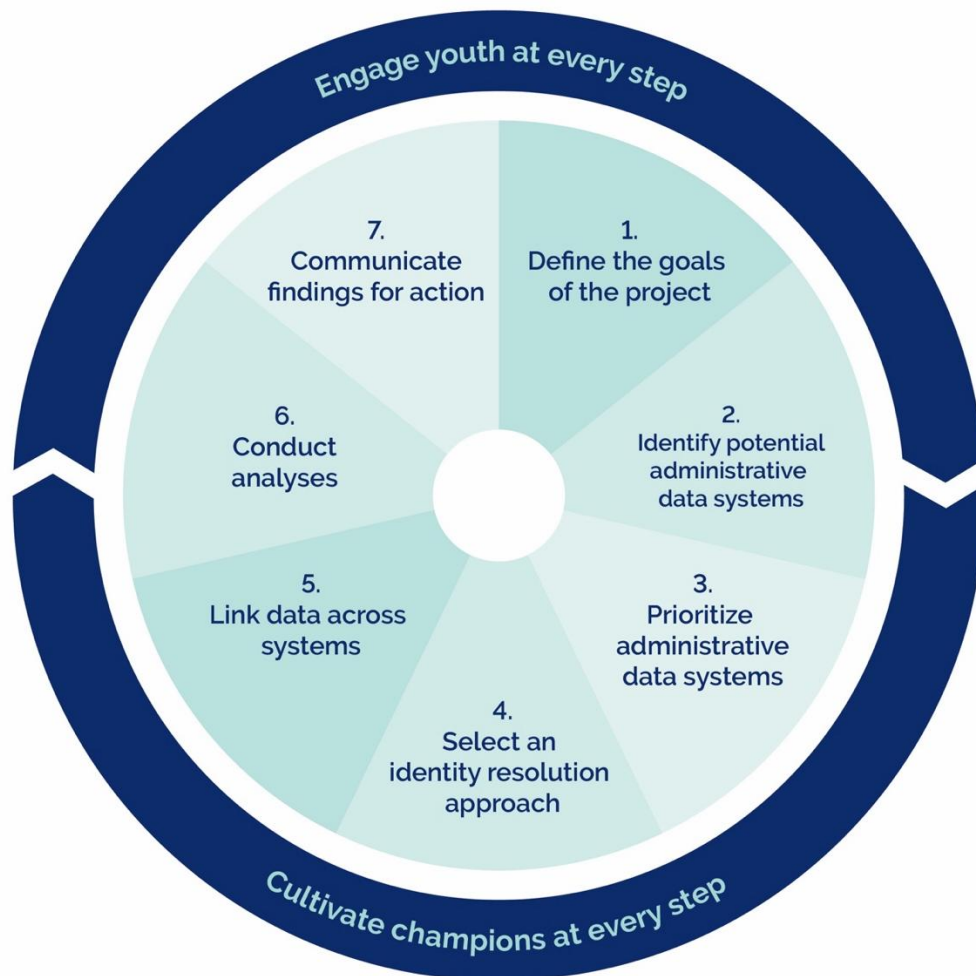
- 1. Define the goals of the study:** What are the ages, characteristics, and geographic areas of interest for estimating youth homelessness? Ideally, what estimates would be generated (e.g., point in time, annual, known counts, unknown estimates)?
- 2. Identify potential administrative data systems:** What are the data systems in those geographic areas that are likely to offer information aligned with the goals of the study? Who provides services to youth experiencing homelessness, and how do they track service delivery? How do youth with lived experience identify themselves in these systems?
- 3. Prioritize administrative data systems:** What are the feasibility and cost considerations for each data system? What does each system uniquely contribute to the estimates or the study goals?
- 4. Select an identity resolution approach:** What identity resolution tools and services are available to connect individual records across systems? Are there existing data sharing agreements or efforts that can be leveraged (e.g., the [Linked Information Network of Colorado](#)'s data sharing agreements)?
- 5. Link data across systems:** What are the common identifiers that allow for identity resolution across the data from each system? Is there a single system contributing to the population of interest or does each system need to contribute a unique portion of the population? What are the inclusion/exclusion criteria that set the boundaries of the individuals included in the data?
- 6. Conduct analyses:** How have the goals of the study evolved? How can the analytic approaches be conducted so they are responsive to the decision-making goals of each system contributing data and the broader landscape of preventing or lessening the duration of youth homelessness? What was learned about the quality, strengths, and limitations of each data source?
- 7. Frame and communicate findings for action:** Who is the target audience? What message and messengers will resonate with distinct audiences? How do you balance reporting priorities given competing interests of key partner agencies?

## Cross Cutting Approaches

The cross-cutting approaches are activities that can be used throughout the iterative approach described above to strengthen the process.

- **Cultivate champions within each system:** Who are the stewards of the administrative data that already are or can become champions for this study? What are their needs and opportunities for use of the study results? Read more about cultivating champions throughout each step of the process in the section titled [Prioritize Administrative Data Systems](#).
- **Engage youth at each step:** How do youth with lived experience report how they appear in systems? Are youth equitably represented in the data systems? How will you ensure youth with various demographic, social, and geographic variations are equitably represented? An in-depth look at how and why engaging youth with lived experience is important to this work is provided in the section titled [Centering the Voices of Lived Experts through an Equity Lens](#).

Figure 1. Process for Linking Cross-Systems Data to Identify and Support Youth Experiencing Homelessness



# Defining Study Goals

Beginning the study with the ANTICIPATED WHY or how the data is initially envisioned to be used is the foundation for a successful study. The WHY will evolve over the design phase of the study to meet the decision-making goals and objectives of each organization contributing data or resources to the study. The WHY helps define:

- What homeless experiences or definitions of homelessness are most relevant?
- What age range or characteristics of youth align with policy or practice decisions?
- What geographic areas are most actionable?
- What timeframes are of interest?
- When would information need to be available to inform policy or practice decisions?
- What types of estimates best align with the goals of the study?

Youth who have experienced homelessness can contribute valuable insights into both the experience of homelessness and homeless services, making their input crucial for defining the questions outlined above. Youth with lived experience should therefore be involved in this initial goal-setting phase to better inform the ANTICIPATED WHY of the study.

## The Power of Linking Administrative Data

Linking administrative data moves estimates of youth homelessness from disparate single system counts to deduplicating counts across multiple systems. Multisystem estimations provide more comprehensive estimates than any single system can produce alone. A multisystem approach to estimating youth homelessness aligns with the practical reality that no one system serves all youth who experience homelessness.

Multisystem estimations can leverage data from a combination of:

- Systems that are designed primarily to serve individuals experiencing homelessness (e.g., Homeless Management Information Services)
- Systems with reporting requirements that are based on identification of youth experiencing homelessness (e.g., State Education Agencies' McKinney Vento Education Data)
- Systems that collect data on homeless experiences for the purpose of delivering a broader suite of services (e.g., Child Welfare).

Linking administrative data across systems is the process of connecting records across systems to learn how many homeless youth are in a geographic area, what services are being used, and what services are needed to support youth who are experiencing homelessness. When data on homeless experiences are linked across systems at the individual level, it is possible to produce more accurate counts of individuals who experience homelessness and have received services in a geographic area and estimate the number of youth who might benefit from services.

## Type of Estimates

In this guide we provide information on two types of prevalence estimates that can be generated by linking data across administrative systems:

*Known Population:* an unduplicated count of youth recognized as experiencing homelessness in a geographic area of interest. Identity resolution across administrative data systems allows for development of contingency tables that summarize how many youth touched each system or multiple systems.

*Unknown Population:* youth estimated to have experienced homelessness by applying a multisystem estimation methodology to the known population data. This approach is commonly used in ecology to estimate the total size of the population and has been applied to public health issues.

These types of estimates can be applied to the population and geographic areas that best meet the study goals. We will discuss known and unknown populations more in depth in the Conduct Analysis section of this Community Guide.

# Identifying Potential Administrative Data Sources

Identifying potential data systems for inclusion in the study can begin by cross walking the study goals with administrative data systems:

1. Defining youth homelessness
2. Considering the contribution of each data source to study goals

## Defining Youth Homelessness

A primary hurdle to counting youth who are homeless is the lack of shared understanding and definition of homelessness across systems that serve these youth, which poses a challenge to generating meaningful data (Cutuli et al., 2019). How you define the target population should align with the goals of the study. There are benefits and challenges to consider when using a broader versus more narrow definition of youth homelessness. A growing consensus within the field of homeless services is that a more sweeping definition could lead to appropriate allocation of resources and service delivery, especially for youth who experience the adverse impact of housing instability but currently fall outside of certain qualifying guidelines (Johnson, 2020).

### Considerations for defining youth homelessness:

- Selection of data systems for use in multisystem estimations begins with defining youth homelessness for the purposes of a study.
- A broad definition of homelessness can be used that encompasses multiple federal definitions of homelessness. Any systems that has data on experiences and aligns with one or more of these definitions can be considered for use in a study.
- Alternatively, a narrower definition might better meet the goals of a community, and then the data systems used must reflect the selected narrower definition.
- Including the voice of youth with lived experience into your definition of homelessness can inform the systems that you use to seek out data. Youth who have experienced homelessness can shed light on their perceptions and how their experiences shape their interactions with systems, thus where and how they appear in administrative data systems.

Table 1 below, provides the definition of homelessness for each system used in the study. It is important to review how these agency definitions overlap and differ. Understanding how the data is collected and defined within these systems is critical to understand the data reported in the systems and how they overlap with one another. Below the agency definitions, the perspectives of youth with lived experience and how they define homelessness sheds light on how youth perceive their experiences.

**Table 1. Definitions of Homelessness**

Source	Definition
<p>Department of Housing and Urban Development</p>	<p><a href="#">The Homeless Emergency Assistance and Rapid Transition to Housing (HEARTH) Act</a> defines homelessness and breaks it into four categories. Below are brief summaries of each category's definition:</p> <ol style="list-style-type: none"> <li>1. Literally Homeless: Individual or family who lacks a fixed, regular, and adequate nighttime residence. This includes people living on the street, in cars, in shelters, in hotels/motels, etc.</li> <li>2. Imminent Risk of Homelessness: Individuals of families who are losing housing in 14 days or less and lack the resources to find subsequent housing.</li> <li>3. Homeless Under Other Federal Statutes: Youth and families who are homeless according to other federal statutes.</li> <li>4. Fleeing/Attempting to Flee Domestic Violence: Individuals or families who are fleeing domestic violence and lack housing or the resources to obtain housing.</li> </ol>
<p>Department of Education - Office of Elementary and Secondary Education - Office of School Support and Accountability</p>	<p>Homeless: Individuals who lack a fixed, regular, and adequate nighttime residence (<a href="#">within the meaning of section 103(a)1</a>); and includes:</p> <ol style="list-style-type: none"> <li>i. Children and youths who are sharing the housing of other persons due to loss of housing, economic hardship, or similar reason; are living in motels, hotels, trailer parks, or camping grounds due to the lack of alternative adequate accommodations; are living in an emergency or transitional shelters; or are abandoned in hospitals;</li> <li>ii. Children and youths who have a primary nighttime residence that is a public or private place not designed for or ordinarily used as a regular sleeping accommodation for human beings (within the meaning of section 103(a)(2)(C));</li> <li>iii. Children and youths who are living in cars, parks, public spaces, abandoned buildings, substandard housing, bus or train stations, or similar settings; and</li> <li>iv. Migratory children (as such term is defined in section 1309 of the Elementary and Secondary Education Act of 1965) who qualify as homeless for the purposes of this subtitle because the children are living in circumstances described in clauses (i) through (iii).</li> </ol>
<p>US Department of Health and Human Services- Office of the Administration for Children and Families- Family and Youth Services Bureau</p>	<p><a href="#">The Runaway and Homeless Youth Act (RHYA)</a> RHYA (42 U.S.C. §5732a) defines homeless youth as</p> <ol style="list-style-type: none"> <li>i. Individuals who are “less than 21 years of age for whom it is not possible to live in a safe environment with a relative and who have no other safe alternative living arrangement.”</li> <li>ii. This definition includes only those youth who are unaccompanied by families or caregivers. This definition is used in connection with the Basic Center Program and the Transitional Living Program.</li> </ol>
<p>How Youth with Lived Experience define homelessness...</p>	<p>“The definition of homelessness is vast. It’s not just sleeping on the street. Homelessness is not only not having a roof, but also not having community, or family, or friends, or a lot of different things [basic needs].” (Homeless youth in Colorado)</p> <p>“Homelessness is suffering. It’s painful. It’s because we have gone through painful things, because people didn’t care for us.” (Homeless youth in Colorado)</p>

## Considering the Contribution of Each Data Source to Study Goals

Cross walking contributions of each data source to study goals is an initial step in study design. At this phase, research teams may use a combination of publicly available data and subject matter experts to begin to rule in and rule out potential data sources for a study.

### Example considerations

- **Age range of youth in the system**
  - Child welfare and education systems can serve young people through age 21; however, most youth exit those systems at age 18.
  - Other systems may serve transitional aged youth and young adults up to age 24.
- **Demographic information**
  - How does the use of data align with available demographic information?
  - LGBTQ youth are at risk for experiencing homelessness. If this population is a priority, is there a data source that includes information on how youth self-identify?
- **Geographic reach**
  - Is information available by the geographic area of interest (e.g., state, county, region)?
  - Some Continuums of Care (CoC) may not offer youth shelters and publicly available data from the local Homeless Management Information System (HMIS) may indicate relatively low numbers of youth served.
  - Geographic information on physical location of youth may not be available in systems that provide remote services (e.g., online schools).
- **Primary purpose of the data system**
  - If the system is designed to serve a particular population of eligible individuals (e.g., only youth with child welfare involvement, only those in public schools), it is important to recognize to whom the results will generalize overall.
  - Most systems are designed to offer services for youth experiencing homelessness who voluntarily (with or without a guardian) seek support. When this is the case, it is important to regularly define the population as youth receiving services from these organizations and not simply youth experiencing homelessness.
  - Because these systems capture information on those seeking support of some kind, this means that there is inevitably an invisible population of youth experiencing homelessness who never touch these systems. This needs to be acknowledged.
- **Capacity to contribute data on a routine basis**
  - If the goal of a study is routine and replicable estimates of youth homelessness, then consider the capacity of each system to routinely provide data.
  - Establishing a process for exporting the same fields, in the same format, on a predictable schedule will help with replication.

## Examples of Administrative Data Systems

Obtaining administrative data for analytic purposes can provide a greater understanding of how youth enter systems and where they access services, leading to more targeted, efficient programs that improve systems for youth. Identifying the universe of system data available will impact what and how you are able to report data. The sources will depend on a combination of factors that include relationships between partner agencies, willingness and ability to share data, and how the purpose of sharing data aligns with the agencies' strategic priorities and goals.

Administrative data systems are primarily housed within government agencies, and most agencies have their own systems. These systems rarely function beyond collecting case specific information for reporting purposes on program specific issues. Looking beyond government agencies, local community service provider agencies may yield more robust and descriptive information that can be merged and analyzed to enhance the quality of each system and what you are able to report.

### Multisystem estimations can leverage data from a combination of:

- Systems that are designed primarily for homeless services (e.g., Homeless Management Information Services).
- Systems with reporting requirements that are based on identification of youth experiencing homelessness (e.g., State Education Agencies' McKinney Vento Education Data).
- Systems that collect data on homeless experiences for the purpose of delivering a broader suite of services (e.g., Child Welfare).

### Examples of administrative data systems include:

For this study, three types of administrative data were linked and analyzed. The administrative data sources used in the study are described first, followed by other potential data sources to consider.

- **Homeless Management Information System:** HMIS includes data from people of all ages accessing services related to homelessness. HMIS is the only national system that collects information on services provided to homeless and unstably housed individuals and families. HMIS is a vital means to identify where homeless youth appear for services and the types of services they receive. The **primary purpose** of HMIS is to allow communities to track services to homeless individuals, allowing for communities to track patterns of where and how individuals access services related to housing instability and to report participation in programs and outcomes for those receiving services.
- **McKinney-Vento Education Data:** McKinney-Vento Education Data: The U.S. Department of Education McKinney-Vento Homeless Assistance Act was enacted in 1987 (and reauthorized under Every Student Succeeds Act, 20 U.S.C. § 6301 (2015)) to address the education of children and youth experiencing homelessness. The McKinney-Vento program ensures the “enrollment, accessibility, and educational stability for students lacking a fixed, regular, and adequate nighttime residence.” The Department of Education requires that each local education agency collects data on youth who are identified for services under the McKinney-Vento program and those data are deduplicated by the state education agency. The **primary purpose** is to identify students who qualify for and receive services under the McKinney-Vento Homeless Assistance Act, a program that provides support for education of youth experiencing homelessness.
- **Child Welfare SACWIS:** Each state has a statewide automated child welfare information system (SACWIS) that collects comprehensive data and serves as a case management tool to support the

administration of child welfare programs in each state. The **primary purpose** of this system is to provide services to children and families involved in the child welfare system. As part of an assessment process, homelessness and risk of homelessness may be documented by case workers. Also, when a young person runs away while in the custody of child welfare, their out-of-home placement status may be documented as “runaway.”

#### Other Potential Administrative Data Sources:

- **Public Housing Authorities (PHA):** Public Housing Authorities are local organizations that receive funding to offer housing assistance, often separate and apart from the organizations that receive funding for emergency shelter and transitional housing support. Partnerships with these organizations would offer additional information about housing instability among youth that would otherwise not be visible in the HMIS partner data. PHA’s may collect data through HMIS or other independent systems.
- **Public benefits programs:** Public benefits programs like the Supplemental Nutrition Assistance Program (SNAP), Child Care Development Funds (CCDF), and Temporary Assistance for Needy Families (TANF) often collect data about factors affecting income and financial stability during the application process, including housing instability. These data partners are worth exploring to see if this information could supplement other data sources.
- **Justice system data including law enforcement systems, department of corrections, and local police data:** It is worth a conversation with law enforcement agencies within the geographic location of interest to understand if and how they may collect information that identifies whether a youth is experiencing homelessness during law enforcement encounters. Some jurisdictions have a process for documenting whether a youth is a runaway or does not have a residential home address, indicating a homelessness experience at the time of the law enforcement encounter.
- **Healthcare claims data:** There is an ICD 10 code for homelessness, Z59.00. However, the consistency of its use may vary across healthcare systems. Healthcare systems that leverage a [Social Health Information Exchange](#) to facilitate closed loop referrals and address social determinants of health may have flags for homeless experiences.

While all the systems described above contain a field that elicits the current and/or past living situation of each youth, none of them are designed to primarily identify youth who may be experiencing homelessness. Leveraging cross-system data can provide a more robust picture of the data on youth homelessness, resulting in a better understanding of how youth enter systems, where youth access systems, and the characteristics of youth. The results can then be used to better target resources to areas where services do not exist or are limited to geographic areas. This results in a more equitable distribution of services to homeless youth.

Identifying systems that hold data the study may want to access is not enough. It is also important to have strong partnerships with local organizations that serve homeless youth, as these connections are vital to not only onboarding data partners, but also to youth who are experiencing homelessness to report on their lived experience. Incorporating the youth voice in reporting sheds light on the youth experience behind the data. Ensuring strong partnerships with agencies and lending the voice of youth with lived experience to the study are not care concepts that should be incorporated throughout the study design. Both concepts are discussed in greater detail throughout this Community Guide.

# Prioritize Administrative Data Systems

Most communities have limited resources and time constraints that make it necessary to prioritize which administrative data systems to use when generating routine and replicable estimates of youth homelessness. Prioritizing which datasets begins with considering the relative contribution of each system to the goals of the study, which were described in the previous section. Then prioritization can become a pragmatic process. What systems have or could have leadership that champion partnering on the study? What systems have the ability to share the unique identifiers necessary to conduct identity resolution? In this section, guidance is provided on how to build the partnerships and feasibility considerations for linking data.

## Cultivate Champions Within Each System

The effort to onboard data systems begins by identifying the partners and organizations who make an important contribution to the goals of the work. Once these groups are identified, the work can only be successful if trusted relationships are built. This means that all groups are dedicated to common goals and see the benefits of the effort to their own organizational missions.

## Identify Systems, Partners, and Primary Contacts at Partner Agencies

Here are some recommended strategies for identifying the teams contributing to the effort:

- **Use an iterative partnership process:** It is rare to have all desired partners involved in the work from day one. Instead, consider a phased approach where the work can move forward with an initial set of partners. This means work can be accomplished while the partners requiring more time to onboard can move at a reasonable pace. Often, the successes of the first phase can jumpstart more skeptical partners to onboard for the next round of work.
- **Prioritize partners:** This iterative partnership approach involves strategically prioritizing the first-phase partners. Recommended priority factors include preexisting relationships with the partners, a known willingness to share data, and the essential nature of their data to the work.
- **Identify primary and secondary contacts:** It is recommended that the first conversations happen with the policy/program leads most interested in the “so what?” of the work being accomplished. This cultivates internal champions at the organization who can then initiate practical data-sharing conversations with their internal legal and privacy teams who need to know that the organization desires to participate in the work.

## Trust-Building Strategies with Organizations and Partners

To help build a successful partnership, it is essential to allow adequate time for the partnership to be developed and formalized. Developing and formalizing the partnership may begin with identifying a common purpose and shared goals (Pasti & Smith, 2019). The collaborative process of identifying a common purpose and goals ensures that all participants come to the table for the shared purpose of addressing these goals. This process helps to establish core values and any potential areas of conflict that may arise. Once the

groundwork for the partnership has been established by identifying common ground among the parties involved, the next step is to create a comprehensive action plan. If the action plan involves working with local CoCs, ensure everyone is brought together to discuss privacy and data sharing. Use existing partnership agreements as templates to help create data-sharing agreements and other shared cooperative agreements. Building relationships and buy-in at all levels of the organization is essential. It also should be assumed that staff turnover will occur. Therefore, focus on long-term sustainability through simple agreements to operate cooperatively and a plan for coordinated actions (Pasti & Smith, 2019).

There are a few strategies that help build trust among organizations considering the partnership effort. These have proven to be successful in Colorado's data sharing work. They include:

- **Identifying value to organization:** A key trust-building strategy is to make sure that the potential partner(s) understands that this is not a one-sided benefit. Once the overall goals of the work are introduced to the new partner, time needs to be spent giving them the floor to describe how they see connections to their own work and decision making. This ensures the effort is not an “off to the side of the desk” item but rather a meaningful asset to their own work.
- **Respecting their capacity:** Often partners have competing demands and are forced to prioritize the effort you are trying to accomplish in a long list of other work. Ask about the capacity of their team to partner in the work, both subject matter experts and data team members. This provides a realistic time frame for their partnership and shows an understanding of their workload and competing job duties.
- **Being flexible:** Rarely does a one-size-fits-all approach work. Provide space for them to express concerns with the data sharing effort and partnership in general. Often, it's an opportunity to think outside the box of how the partnership was originally envisioned and allows for flexibility depending on the needs of the partner.
- **Addressing benefits and risks:** Once a partner has made their own connections about how the work is a benefit to them, it's important to not ignore potential risks or burdens of the work. This helps the new partner understand that their sacrifices are not invisible and is also an opportunity to shape solutions that reduce the potential risks or dispel mythical risks that don't really exist.
- **Supporting their legal and privacy needs:** Often, the partner organizations are not familiar with all the privacy laws and regulations related to data sharing. If they don't have the in-house expertise to know what's allowed and what's not, help them find a resource, such as federal guidance on data sharing for research and evaluation that is specific to their system (e.g., The [HIPAA Privacy Rule](#)) that will give them the confidence to pursue the partnership under allowable conditions.
- **Phoning a friend:** If there is a similar partner already involved, be strategic and ask them to join the initial conversations with new partners so there is the comfort of “I'm not the first one.”
- **Deepening partnership over time:** Build relationships and buy-in at all levels of the organization. Assume staff turnover and focus on long-term sustainability. Allow adequate time for partnerships to be developed and formalized.
- **Not reinventing the wheel:** Borrow and replicate [processes](#) and [agreements](#) that have been used with other partners. This helps create institutional models for your work, but also streamlines and creates efficiencies in the onboarding process for partners.

## Practical Feasibility of Linking Data from Each System

All the work described above is only worthwhile if there are not “fatal flaws” that prevent the partner’s data from being included in the linked dataset. Some key considerations that need to be addressed quickly in the partnership include:

- **Data privacy restrictions:** Though uncommon, there are occasionally partners who have explicit laws or regulations that restrict what data can be included in this effort. If this is the case, it’s important to identify that up front so expectations are set appropriately.
- **Common identifiers across data sources:** For this work to be successful, there needs to be common identifiers used for linkages across the data sources (e.g., name, date of birth, social security number, etc.). There are sometimes partners who can serve as the “Rosetta stone” between two data sources with uncommon identifiers, such as Driver’s License Records, but this is rare and can be difficult to implement. If you don’t have enough common identifiers across sources to allow for adequate linkages, this may prove to be insurmountable.
- **Centralized nature of data:** When data that are collected locally are centralized at the regional or state level, exporting those data can be more efficient. In the Companion Study, data sharing agreements and data exports were feasible at the state level for education and child welfare information because all data were available in a statewide administrative data system. The HMIS data were housed regionally and required all regions to approve data sharing for the project. The export was efficient because one data analyst was able to export data from all regions through centralized data access, reducing the burden on other regions.

# Prepare and Link Data

Once the partnerships are in place, there are some key considerations during the data sharing process. These include:

- **Design the study to meet purpose:** What are the primary goals of the work? Who will take action with the results? To what end and how often? What are the specific questions we want to answer to meet the goals? These are the important questions to answer before the data sharing model is established because this will influence the scope of data as well as the frequency data need to be shared. If it's a one-time study, simple data extracts being linked manually makes sense. If this is something that's desired to be updated routinely, then an automated linkage and analysis model may be most appropriate. Defining these features will help guide the data sharing approach.
- **Confirm common identifiers:** While this will be done at a high-level during step 3 (i.e., assessing the practical feasibility of linking data from each system), it will be important to understand more specifically what these identifiers look like in the various data sources, how they're collected and for what purpose. This often reveals which identifiers are most reliable and which have vulnerabilities that require intentionality to address.
- **Become a close colleague with the data expert:** The resulting work is only as good as the data being used, and no one knows the data better than the person who works most closely with it every day. The best person is the one who understands the programmatic processes for data coming into the system (who enters the data? how does it get entered? what quirks exist in this process? when did fields change over time?) and uses the data regularly for reporting purposes. They'll have the best intelligence to offer so data are not unintentionally misused or misinterpreted in the work.
- **Set regular quality checks:** Checking the quality of the work at every stage promotes data quality. Any data analyst can attest that human error is inevitable in coding, data cleaning, data linkage, and analysis. The goal is not necessarily to get it entirely right the first time around (though that would be nice). The goal is to have a safety plan to catch errors when they occur and correct them before it's too late. In this type of data work, it can have drastic consequences if a misstep is taken (accidentally dropping an entire group of individuals from the data, missing matches that should have been made). Ideally, there are at least two analysts who can check one another's work along the way to ensure the highest quality possible.
- **Select an identity resolution strategy:** While there is not a specific gold standard for identity resolution, there are many approaches one could take. The first factor that influences this decision is whether this is a one-off study or a routine study. If it is routine, it is recommended that the team identify an automated solution to the linkages that can be easily and efficiently accomplished. Though this often results in more errors (no hand curating of the match results is done), it meets timeliness goals and the error may be acceptable. If this is a one-off or less routine (e.g., twice a year), a manual approach may be ideal. Often, this includes statistical coding using a rule-based matching strategy (e.g., social security numbers are an exact match), followed by more "relaxed" or fuzzy matching to pick up matches where human error occurred during data entry (e.g., first name switched with last name, date of birth year is transposed). Sometimes, with very small populations (youth experiencing homelessness may qualify), it is worth the time to look for individuals one by one in the data to avoid big swings in results. Though time intensive, it may be worth it if the study is only being accomplished once or twice.

For more information on the process used for linking administrative data in this study, read about [LINC in the Appendix](#).

# Conduct Analysis

## Analytic Decisions

Conducting analysis using administrative data often requires a series of analytic decisions, each of which must be well documented to ensure replicability of findings and accurate explanation of the results. Key decision points need to be made around the following data items:

**Demographic data:** information on individuals age, gender, and race/ethnicity may not align across systems. Considerations might include:

- How data are collected (e.g., individual self-identification)
- If data are validated
- Categories of data collected (e.g., federal reporting categories, more nuanced information)
- Amount of missing data
- At what level data are deduplicated by a system (e.g., local education agency, state education agency)

**Geographic data:** State-level estimates require fewer analytic decisions than analyses at local geographical areas but may be less actionable in guiding investments. Consider how to treat data when:

### *Examples from the Study*

- service areas do not directly overlap across systems (e.g., county and school districts)
- geographic data reported within a system are inconsistent. For example, in the HMIS data, a few geographic areas were reported as a city rather than a county. To match on county across all three data systems, we assigned a county when geographic information was provided at the city level. In all cases, the cities did not exactly map one-to-one to a single county. However, the vast majority of each city's population did live in one county, so we chose that county as the designated county for that city.
- geographic data are missing or not reflective of the youth's physical location. For example, data from Colorado Department of Education for students participating in a virtual school learning environment. We did not attempt to assign a county location to these youth. These youth can be included in the statewide analysis but would get dropped from a geographic-specific analysis.

**Annual versus Point in Time Estimates:** Annual estimates may be more feasible when using data that are linked across systems.

- Some systems may not have dates when a homeless episode began or ended, eligibility for a service might be tied to a school year or calendar year.
- Point in Time when homelessness was measured might not be on the same day across systems; thus, a fall education "count" might not sync up with a winter "point in time survey."

## Types of Estimates: Known and Unknown

**Known Population:** an unduplicated count of youth recognized as experiencing homelessness in a geographic area of interest. Identity resolution across administrative data systems allows for development of contingency tables that summarize how many youth touched each system or multiple systems.

### Case Study

Sarah is a 17-year-old youth who uses she/her pronouns who has experienced homelessness after running away from home multiple times. Sarah spent several years in the foster care system in Colorado before being adopted but has a strained relationship with her adoptive mother and does not feel safe or comfortable at home. Although she receives homeless services to help meet her basic needs, she is not formally enrolled in a program and is only served on a drop-in basis.

Because Sarah is a minor, she would need parental permission to be housed through her local RHY provider and her adoptive mother refuses to provide her with permission. Instead, Sarah regularly couch surfs, sleeps in her car, and relies on her local RHY provider to meet her basic needs.

Sarah has struggled with her mental health and because of the lack of services available in rural Colorado, she had to be taken by ambulance to Denver to receive the help she needed. She has also had some run-ins with police, who typically just return her to her adoptive mother’s house even though she has communicated that she does not feel safe there. Although she has had multiple interactions with the police, she has never been arrested and would not appear in the local police’s database. Despite her struggle with homelessness, Sarah has never been flagged by the education system and has not shared any information with teachers or counselors at school. With the critical support of her local RHY provider, Sarah is on track to graduate high school in a few months and plans to go on to college afterwards.

Table 3. Known to a System\*

Known to a System							
	Education	Child Welfare	Homeless Services	Education & Child Welfare	Education & Homeless Services	Child Welfare & Homeless Services	All Three Systems
Sarah	N	Y	Y	N	Y	Y	N
Colorado Youth 14-17 in Fiscal Year 2022							

\*Based on details provided by youth and not on review of actual administrative records.

**Unknown Population:** youth estimated to have experienced homelessness by applying a methodology to the known population data that is commonly used in ecology to estimate the total size of the population. The multisystem estimation method, commonly referred to in ecology as the “capture- recapture” approach, begins by setting a timeframe of interest, then linking individuals across the systems to determine if a youth was recognized as experiencing homelessness in a single system or multiple systems.

Statistical models are applied to contingency tables to produce an estimate of the unknown population. This process has been employed to estimate marginalized and difficult to measure populations such as those with problematic drug use, opioid use disorder, and undetected COVID-19 infection.

In the study, we used this methodology to generate a rigorous estimate of the total population of youth ages 14-17 in Colorado who experienced homelessness.

[Figure 2 from the Full Report](#)

[Link to the Full Report](#)

Figure 2.



**Total Population:** the sum of the known population and the unknown estimate is a rigorous multi-system estimate of the total population of youth experiencing homelessness.

## Incorporating Youth with Lived Experience in Data Analysis

Giving voice to the youth with lived experience behind the administrative data is critical to understanding the challenges, barriers, and systemic issues faced by youth experiencing homelessness. Linking and analyzing data collected in administrative systems and presenting counts of the number of unstably housed youth, and where and how they access services helps to shed light on the problem by providing a voice to those youth behind the numbers, which can lead to improved systems.

To highlight how youth appear as known and unknown in administrative data systems, the following case study presents the story of “Grant,” a youth in rural Colorado that the research team spoke with about his experience as a youth that led to housing instability.

### Case Study

Grant\* is a 21-year-old youth who uses he/ him pronouns and who has experienced homelessness and housing instability since turning 18. Grant’s mom has not been part of his life, and his dad struggled with substance abuse and had a difficult time raising him on his own. As a result, Grant spent about 2 years in foster care in another state beginning around age 12. When he was 13, his grandparents in another state took custody of Grant. For a few years, he bounced between living with his dad and his grandparents, but after many disagreements with his grandparents, he decided to live with his dad again at age 16, which brought him to a small mountain town in Colorado. Grant started to struggle

with substance abuse around age 18, he was kicked out of his dad’s house and became homeless, couch surfing and sleeping in parks as he moved around. Many times, Grant encountered the police, but most interactions were due to his sleeping in public spaces and he was never jailed. He has spent the past 3+ years couch surfing across the country between his grandparents’ home and his dad’s home in rural Colorado. Grant said the only way he can survive is with services provided by the local Runaway and Homeless Youth Services (RHY) provider.

Despite his lack of safe and stable housing, Grant shows a great deal of resilience. With the support from his RHY provider, he received his high school diploma and is employed full-time. Additionally, the local RHY provider helped Grant secure vital documents like his driver’s license and helped him to open a checking account at a local bank; he is currently saving money. Grant is making plans to move into his own rental apartment in the next year or so. With the support of a local RHY provider, Grant is confident in his ability to secure long-term safe and stable housing and maintain full time employment. Without their help, Grant said he would be on the streets, probably in Denver, as he could not stay in the rural community where he currently lives with no services or support. As he put it, “I think another big part about [RHY provider] is less the financial support and more just like the social aspect of it. Like actually having somebody to talk to, to organize all these things and develop plans for the future That’s also really helpful too.”

**Table 2: Factors influencing Grant’s KNOWN and UNKNOWN appearance in systems**

Administrative Data System	Factors that determine if youth is identified in administrative data system
<b>Child Welfare (foster care) SACWIS</b>	<ul style="list-style-type: none"> <li>State/location received services from child welfare.</li> </ul> <p><i>Grant is not identified in the Colorado SACWIS system because he was not in foster care in Colorado.</i></p>
<b>Homeless Management Information System (HMIS)</b>	<ul style="list-style-type: none"> <li>Where he received support from RHY provider.</li> </ul> <p><i>Grant appears in HMIS because he receives services from the local RHY provider</i></p>
<b>McKinney Vento Homeless Assistance Act (Education data)</b>	<ul style="list-style-type: none"> <li>Age and location of school that youth attended where he was identified as experiencing homelessness.</li> </ul> <p><i>Grant is unknown in the McKinney Vento data because it is undetermined if he was connected to McKinney Vento services while attending school in Colorado.</i></p>
<b>Police Department</b>	<ul style="list-style-type: none"> <li>Location/state he received a trespass ticket for sleeping in a park.</li> </ul> <p><i>Grant is unknown in police department data. It is unknown if he was ticketed in Colorado.</i></p>
<b>Medical/Hospital records</b>	<ul style="list-style-type: none"> <li>Location/state he went to hospital/ ER for substance use treatment.</li> </ul> <p><i>Grant would only be known in medical/hospital records if he received treatment in Colorado and a provider included an ICD Code for homelessness, Z59.0, or other flag to indicate homelessness.</i></p>

\*The name of the homeless youth has been changed to protect their privacy and ensure confidentiality.

# Centering the Voices of Lived Experts Through an Equity Lens

Over the past several years, there has been a shift towards incorporating the voice of youth with lived experience into policy and programming by government agencies, researchers, and service providers. A more concentrated effort has been made to engage with the populations being served to learn from youth with lived experience to provide more targeted and responsive programming and policies. A 2024 U.S. Department of Health and Human Services, Administration for Children and Families Information Memorandum on “Leading in Partnership with Youth and Young Adults with Lived Experience” provides guidance to researchers, practitioners, and policymakers on ways to meaningfully partner and collaborate with youth with lived experiences (ACF-ACYF-FYSB-IM24-01, 2024). This guidance was incorporated in the framework for engaging with youth with lived experience in the research design, understanding the story behind the data, and interpreting results.

Engaging with youth with lived experience who are part of a vulnerable population, such as youth experiencing unstable housing and homelessness, requires careful planning and consideration. Engaging with youth with lived experience to hear their stories provides a clearer picture of how these youth appear in systems and show up in the data. Consider holding a focus group to generate input and feedback to the data request. Once the analysis is complete, convene youth with lived experience again to hear their stories to help put the data in context. Engage with youth to help define the approach to analysis and then describe the results.

To engage with youth, work closely with partners engaged in the work; in this case, we reached out to runaway and homeless youth service providers to help coordinate interviews and focus groups with youth. Ensure youth have a safe, comfortable space to share their stories. Approach questions with curiosity and take care with your words. Gain informed consent and ensure youth of their confidentiality and the voluntary nature of responding to questions. Make sure they do not feel pressured to answer any questions and let them know it is okay to not answer any question they are not comfortable answering and share the experiences they want to share. Provide food, snacks, transportation, and payment or an incentive as a thank you for sharing their time and experiences with you. Thank youth for their participation and offer space to talk one on one after or refer to support if needed following discussion. Support the voice of youth with lived experience through an equity lens. Consider talking with youth from diverse audiences including race, age, system involvement, access to services, geography (rural versus urban versus suburban youth all have different experiences). Incorporate the voices of all youth with diverse backgrounds to ensure their perspectives are heard.

Figure 3. Process for Engaging Youth With Lived Experience



# Frame the Message and Communicate Findings

Early in the process of framing the message and communicating the findings, it is vital to share any initial results with primary data partners to garner feedback. This will create an opportunity to demonstrate where the gaps may be and where data may be missing. When framing the message from these data, it is also important to demonstrate how any existing counts generally underestimate youth homelessness (Auerswald & Adams, 2018). To help further contextualize the quantitative findings and highlight the adverse experiences of homeless youth, incorporate the lived experiences of these youth using qualitative data (Padwa et al., 2023). Presenting results in accessible formats, such as infographics, presentations, and short briefs, can promote actionable use of the data. Regardless of the format, the results should be tailored to reach various audiences including partners and community members.

- Incorporate qualitative information from youth with lived experience to help provide context to quantitative findings
- Share initial analysis with primary data partners
- Use as opportunity to demonstrate where the gaps may be, where data may be missing
- Demonstrate how existing counts are generally an undercount
- Report results in an accessible format
- Tailor to wide variety of audiences
- Target specific agencies

Examples of target audiences include youth with lived experience, champions, partner agencies that provide data, and practitioners, policymakers, and researchers in the broader community. An approach to framing should be a step process in which each target audience helps to inform the framing of the message moving forward. Findings should first be communicated with youth with lived experience for input and feedback. Youth with lived experience would then inform the framing of the message to study champions and their organization more broadly. The feedback from the youth with lived experience, study champions, organizations with the administrative data would then inform the message given to the broader community. This ensures that a broad audience understands the interpretation of the data and the context and implications to the community.

# Conclusion

Leveraging administrative data to identify youth experiencing homelessness is a powerful tool in preventing and ending youth homelessness in targeted geographic areas. Identifying youth experiencing homelessness in, and across systems allows communities to learn about the prevalence of youth homelessness, risk and protective factors, and services youth access. This in turn allows for a better, more targeted allocation of resources across agency and geographic area. By listening to youth who are experiencing homelessness, and how they access services, practitioners can better understand the needs of the population and target resources to fill gaps in services.

This Community Guide provides a framework for communities to use when linking administrative data is possible to better identify, understand and pinpoint resources and services for youth experiencing homelessness. The seven iterative steps and cross-cutting approaches are an overarching framework using linked administrative data to inform, prevent and lessen the duration of youth homelessness. Among the most important steps is defining the goals of the study - and selecting a definition of youth homelessness that is aligned with the study goals. Communities using this guide can determine if taking a broad definition of youth homelessness when accessing administrative data will yield more robust information (e.g., population estimates) to achieve study goals or if a narrower definition will lead to a more targeted activation of the findings (e.g., resource allocation with eligibility restrictions on funding). We encourage communities using this guide to continually return to the goal of their study and use the goal of the study to guide subsequent decisions. Taking this iterative approach to study design and incorporating the voice of lived experts will lead to a well-designed study that yields strong, actionable results.

# References

- Administration for Children and Families. (2024). ACF-ACYF-FYSB-IM24-01, 2024. Retrieved February 20, 2024, from <https://www.acf.hhs.gov/fysb/policy-guidance/fysb-information-memorandum-young-adults-lived-experience-february-2024>
- Auerswald, C. L., & Adams, S. (2018). Counting all homeless youth today so we may no longer need to tomorrow. *Journal of Adolescent Health*, 62(1), 1-2. <https://doi.org/10.1016/j.jadohealth.2017.10.013>
- Barocas, J. A., White, L. F., Wang, J., et al. (2018). Estimated Prevalence of Opioid Use Disorder in Massachusetts, 2011-2015: A Capture-Recapture Analysis. *Am. J. Public Health*. 2018;108(12):1675-1681. doi:10.2105/AJPH.2018.304673
- Böhning, D., Rocchetti, I., Maruotti, A., & Holling, H. (2020). Estimating the undetected infections in the Covid-19 outbreak by harnessing capture-recapture methods. *Int. J. Infect. Dis.* 2020;97:197-201. doi:10.1016/j.ijid.2020.06.009
- Boullion, A., Holland, M.L., Brock, S.E. et al. Rural Families Experiencing Homelessness: Demographics, Program Design, Outcome of High Barrier Transitional Housing and Rehabilitation Program. *J Child Fam Stud* 31, 1823-1837 (2022). <https://doi.org/10.1007/s10826-021-02144-6>
- Cutuli, J. J., Treglia, D., & Herbers, J. E. (2019). Adolescent homelessness and associated features: Prevalence and risk across eight states. *Child Psychiatry & Human Development*, 51(1), 48-58. <https://doi.org/10.1007/s10578-019-00909-1>
- Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). <https://www.congress.gov/bill/114th-congress/senate-bill/1177>
- Gemmell, I., Millar, T., & Hay, G. (2004). Capture-recapture estimates of problem drug use and the use of simulation based confidence intervals in a stratified analysis. *J. Epidemiol. Community Health*. 2004;58(9):758-765. doi:10.1136/2003.008755
- Johnson, L. A. (2020). A consistent definition of homelessness in ACF programs. *Administration for Children and Families*. Retrieved from [www.acf.hhs.gov](http://www.acf.hhs.gov)
- Min, J. E., Pearce, L. A., Homayra, F., et al. (2020). Estimates of opioid use disorder prevalence from a regression-based multi-sample stratified capture-recapture analysis. *Drug Alcohol Depend.* 2020;217:108337. doi:10.1016/j.drugalcdep.2020.108337
- Olivier, J., & Russell, J. (2023). *Authentic youth engagement within organizations: What does it look like in practice?* Chapin Hall at the University of Chicago. [https://www.chapinhall.org/wp-content/uploads/Chapin-Hall\\_Authentic-Youth-Engagement-Study\\_Brief-Sept-2023-1.pdf](https://www.chapinhall.org/wp-content/uploads/Chapin-Hall_Authentic-Youth-Engagement-Study_Brief-Sept-2023-1.pdf)
- Padwa, H., Henwood, B. F., Ijadi-Maghsoodi, R., Tran-Smith, B., Darby, A., Bluthenthal, R., ... & Gelberg, L. (2023). Bringing lived experience to research on health and homelessness: Perspectives of researchers and lived experience partners. *Community Mental Health Journal*, 59(7), 1235-1242. <https://doi.org/10.1007/s10597-023-01138-6>
- Pasti, L., & Smith, T. (2019). Identifying and Addressing Barriers to Program Implementation: A tool for Community Reflection (pp.1-24). <https://youth.gov/sites/default/files/Implementation-Barriers-Tool-JFF-FYI.pdf>
- Thompson, K., Barocas, J. A., Delcher, C., et al. (2023). The prevalence of opioid use disorder in Kentucky's counties: A two-year multi-sample capture-recapture analysis. *Drug and Alcohol Dependence*. 2023;242:109710. doi:10.1016/j.drugalcdep.2022.109710
- Troisi, C. L., D'Andrea, R., Grier, G., & Williams, S. (2015). Enhanced Methodologies to Enumerate Persons Experiencing Homelessness in a Large Urban Area. *Evaluation Review*, 39(5), 480-500. <https://doi.org/10.1177/0193841X15610191>

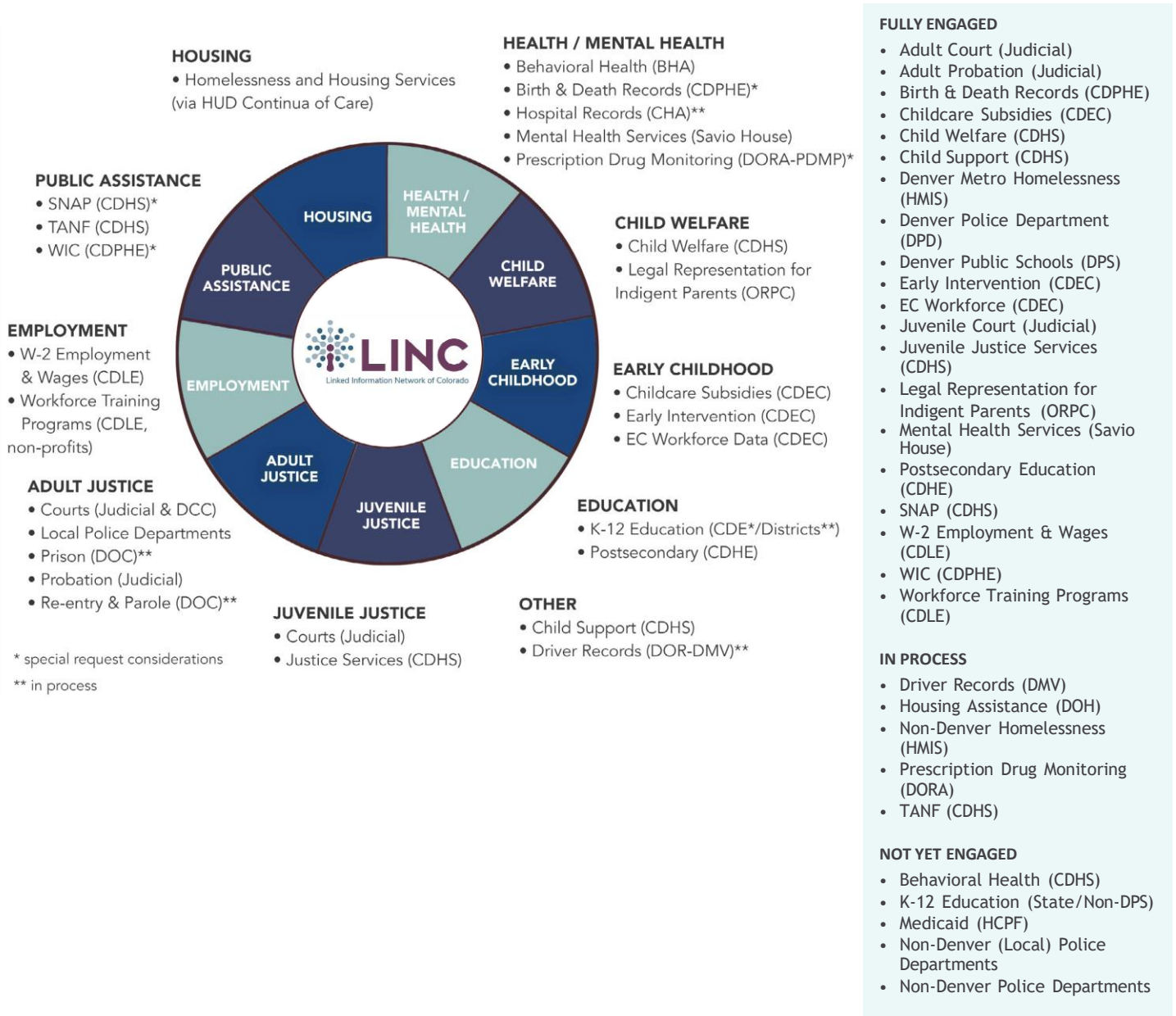
# Appendix

## What is LINC?

### Linked Information Network of Colorado (LINC)

Public systems can better address pressing needs while saving money by using data to understand the complex challenges people face. Everyone interacts with public systems in different ways.

Looking at data from one agency doesn't tell us enough to improve the lives of Coloradans. We need to integrate information across organizational silos to address problems holistically. LINC securely connects data across systems and removes identifiable information so it can be used to make better decisions while protecting privacy. LINC bridges the chasm between the potential of data and its actual use for research and analytics to help solve societal problems in a cost-effective way.



## Benefits of LINC Collaborative

- **Secure, quality data.** LINC delivers the highest available standards to ensure data privacy, and its expertise in top-tier identity resolution provides customized data matching capabilities.
- **Big data, tailored insights.** The large and growing collaborative of data partners offers unique cross-system insights on real-life challenges along with a tailored approach for each study.
- **Time and cost savings.** New partners are compensated for their time to contribute data to a study, and partners have the benefit of a time-saving single data request pathway.

## How LINC Works

### Project Development

LINC projects are born from real problems government partners face that demand analytics to inform smart decisions. Frequently these project ideas arise during government leadership meetings, task force convenings, or informal brainstorm sessions. Researchers can also initiate LINC projects if they have expertise to contribute to an area of high priority to government partners.

### Data Partnering

State government agencies benefit from LINC's streamlined, secure process to research complex policy, process, and service issues. The LINC Director supports LINC project requests so they are relevant and likely to be successful. LINC data partners have an opportunity to provide feedback and ultimately make the decision on whether the project moves forward.

### Link and De-Identify Data

LINC is not a data warehouse. It is a federated data model where data partners temporarily provide data for an approved LINC project. LINC follows all data security and privacy standards required by state and federal laws and regulations. The LINC Data Scientist secures the LINC project data in an encrypted environment and performs a series of identity resolution strategies. The LINC Data Scientist then de-identifies the linked project data to meet the standards of all LINC data partners.

### Ethical Data Use

The LINC Director provides continued oversight of LINC projects once the de-identified LINC project data are in the hands of the approved requester. This ensures the LINC project data are only used for the approved purposes outlined in the LINC Data Use License (DUL) signed by the requester's organization. The LINC data partners can review and share feedback before any results are shared publicly. The LINC requester must destroy the de-identified LINC project data once the project period is over.

## Data Approach for the Linked Information Network of Colorado (LINC)

The following is the typical sequence of the data work for a LINC project as an example of what could be done for studies not requiring automated work:

1. A project request is submitted by the analyst team to the LINC management staff. This request is reviewed by all data partners whose data are being requested.
2. A virtual meeting is held to improve the request with data partner feedback. This involves clearly defining the common identifiers, variables of interest, and contextual information about these data necessary for the data linking and analysis to be accurate.

3. A final version of the request is reviewed and approved by the data partners.
4. Data are extracted by data partners and shared through a Secure File Transfer Protocol (SFTP) to the LINC analyst computing environment.
5. LINC analysts compare the data received to the approved request to ensure all fields were received and are populated in the data. If not, the LINC analyst follows up with the data provider to rectify the data.
6. LINC analysts apply the inclusion criteria for each data source so the data being linked are inclusive of those relevant to the work.
7. The following steps are taken by a primary LINC analyst and all code and resulting data are reviewed by a secondary LINC analyst for quality assurance:
  - Deduplicate records before linking, resulting in a single record per unique person in the data.
  - Standardize common identifiers so they have the best chance of matching if they are the same person.
  - Perform identity resolution using first a rule-based, then a relaxed, approach.
  - Quality review of the matches which involves randomly sampling and checking for false positive and false negative matches. These serve as the match rate estimates.
  - Pull in original list of key variables from each data source based on the approved request.
  - De-identify data to the standards required for all LINC projects.
  - Produce a data match report with the data dictionary for data partners to review before sending to approved requestor.
  - Send out data for analysis. Deduplicate records before linking, resulting in a single record per unique person in the data.