Using Administrative Data to Understand Youth Homelessness: A Data Mapping Guide

Youth Homelessness Data Solutions Project

Madeline Youngren | Amy Dworsky   **March 2024**

Introduction

In 2020, Chapin Hall received an award from the U.S. Department of Housing and Urban Development (HUD) for the Youth Homelessness Data Solutions Project (YHDSP). YHDSP is a collaboration between Chapin Hall and Partnership Center Limited along with Continuums of Care (CoC) in Travis County, Texas; New York City; and the Midlands Area, South Carolina. The project's objective is to develop innovative methods for counting youth experiencing homelessness, estimating the prevalence and incidence of youth experiencing homelessness, generating information about the characteristics of youth experiencing homelessness, and exploring pathways into homelessness using Homelessness Management Information System (HMIS) data linked with administrative data from other sources.

Chapin Hall developed this data mapping guide for communities that want to leverage their HMIS data paired with other administrative data to learn more about their population of youth experiencing homelessness. The guide explains how communities can identify, catalogue, access, and assess potentially relevant administrative data. It also provides useful tools to help with each of those steps. The guide was informed by the experiences of the YHDSP project team, our partner CoCs, and other experts.

Data Mapping

WHAT IS DATA MAPPING?

We conceptualize data mapping as a four-step process: (1) identifying relevant data sources, (2) cataloguing the data elements available from each source, (3) accessing individual-level data from each source, and (4) assessing data quality (see Figure 1). Each of these steps is described in more detail below.

Figure 1. Steps to Data Mapping

Identify Data Sources — STEP 01

Catalogue Data — STEP 02

Access Data — STEP 03

Assess Data Quality — STEP 04
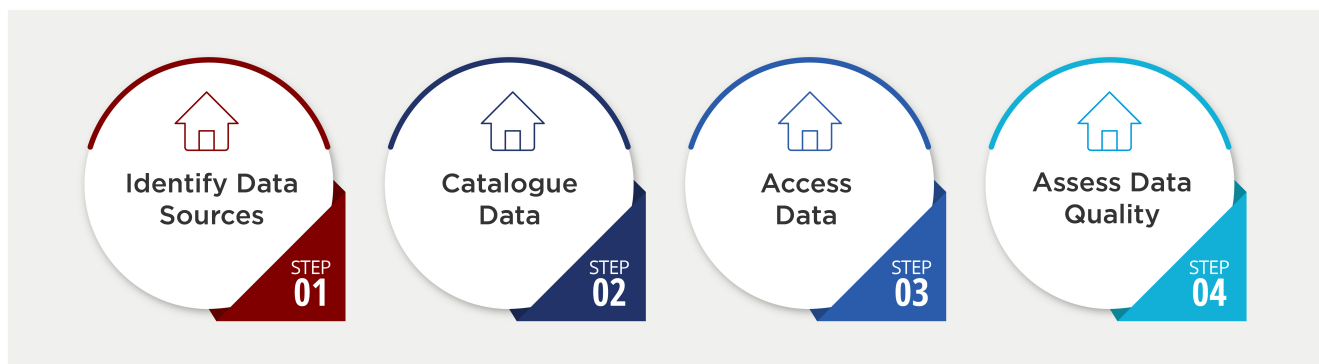
WHY MAP DATA?

Data mapping can provide a blueprint for using administrative data to understand the characteristics of youth experiencing homelessness and young people's pathways into homelessness through other systems. The use of administrative data can be more efficient and better informed when data have been conceptually mapped. Data mapping can also reveal potentially overlooked data sources.

Figure 2. Data Mapping Can Reveal Potentially Overlooked Data Sources
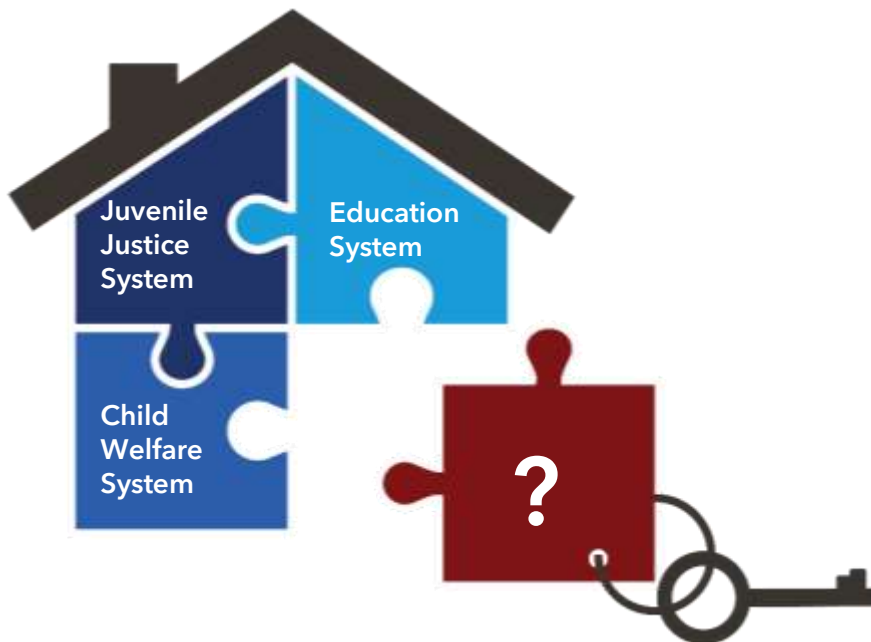
Data Mapping Process

IDENTIFYING DATA SOURCES

Data mapping starts with identifying sources of administrative data that can be combined with HMIS data to (1) generate information about the number and characteristics of youth experiencing homelessness, (2) explore young people's pathways into homelessness through other systems, and (3) identify predictors of homelessness. These data sources include all the systems or agencies youth experiencing or at risk for homelessness interact with as well as other potentially relevant sources of data. A structural relationship map and a case flow diagram can help identify and make sense of these data sources.

Structural Relationship Map

A structural relationship map visually represents the systems and agencies with which youth experiencing or at risk for homelessness interact. A structural relationship map may also include other potentially relevant sources of data, including data from nongovernmental sources (for example, nonprofit organizations that provide services to youth experiencing homelessness without state or federal funds). The map can illustrate relationships between entities as well as between entities and their data. Data can be mapped at national, state, or local levels. Table 1 lists some of the data sources that a structural relationship map might include.

Table 1. Potential Data Sources for a Structural Relationship Map

Source	Description
Homelessness Services System or Continuum of Care	CoCs operate an HMIS to collect client-level data and data on the provision of housing and services to individuals and families experiencing or at risk for homelessness. HMIS data can be used to count the number of youth experiencing homelessness, describe their characteristics, and produce prevalence and incidence estimates (when combined with Census data). However, HMIS data provide an incomplete picture of youth homelessness. Youth experiencing homelessness may choose not to access homeless services, may not be eligible for those services because they do not meet HUD's definition of homelessness (which, for example, does not include couch surfing or living doubled up), or may live in a community where services do not exist.
Education System	Under the federal McKinney-Vento Homeless Assistance Act, schools are required to track which students experience homelessness during the school year and whether students experiencing homeless are unaccompanied (that is, not living with a parent or legal guardian). Additionally, because the McKinney-Vento Homeless Assistance Act defines homelessness more broadly than HUD, school data capture young people experiencing homelessness who would be excluded from HMIS because they are couch surfing or living doubled up. Hence, they can be used to supplement counts and prevalence or incidence estimates based on HMIS data alone. School data can also be used to identify predictors of youth experiencing homelessness such as dropping out.[1]
Child Welfare System	Child welfare system data may include data on child maltreatment investigations, placement prevention services, and foster care placement histories. These data are important because youth in foster care are at high risk for experiencing homelessness after they exit the child welfare system, especially by aging out.[2-6]
Juvenile Justice or Criminal Legal System	Juvenile justice or criminal legal system data may include data on offenses committed by youth, the disposition of charges, and detention or incarceration. These data are important because youth with a history of justice system involvement are over-represented among youth experiencing homelessness.[7-10]
Mental and Behavioral Health System	Mental and behavioral health system data contain information about receipt of outpatient mental and behavioral health services. These data are important because youth experiencing homelessness exhibit high rates of substance use and mental health disorders, such as depression, anxiety, and posttraumatic stress disorder.[11-13]
Vital Statistics	Vital statistics data include data on births and deaths. These data are important because youth experiencing homelessness have a much higher mortality rate than their peers in the general population and many youth experiencing homelessness are pregnant or parenting.[14,15]
Medicaid	Medicaid enrollment data can be used to determine whether young people experiencing homelessness are enrolled in their state's Medicaid program. Medicaid claims data can be used to examine the health conditions for which young people experiencing homelessness have received treatment. This is important because youth experiencing homelessness face barriers to accessing healthcare and are at higher risk for a range of negative physical, mental, and behavioral health outcomes.[16-18]

Figure 3 illustrates what a structural relationship map might look like. The contents of the map will vary depending on the state and local context.

**Figure 3. Generalized Structural Relationship Map**
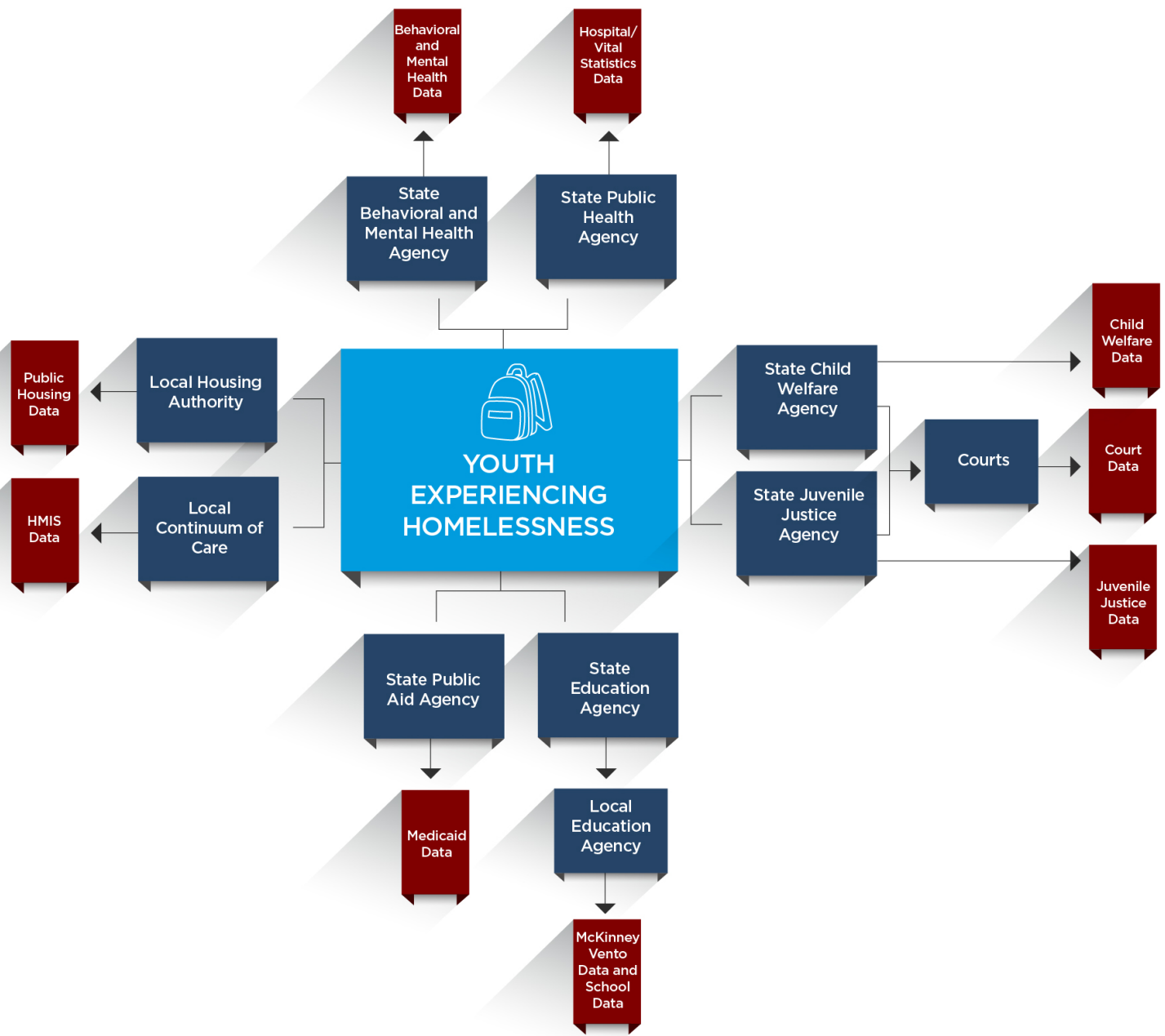
A case flow chart visually represents all the possible sequences of systems through which youth at risk for homelessness might enter the homelessness system. Creating a case flow chart can help identify potentially relevant data sources. Figure 4 shows just a few of the many possible sequences.

Figure 4. Youth May Enter the Homelessness System Through Various Pathways
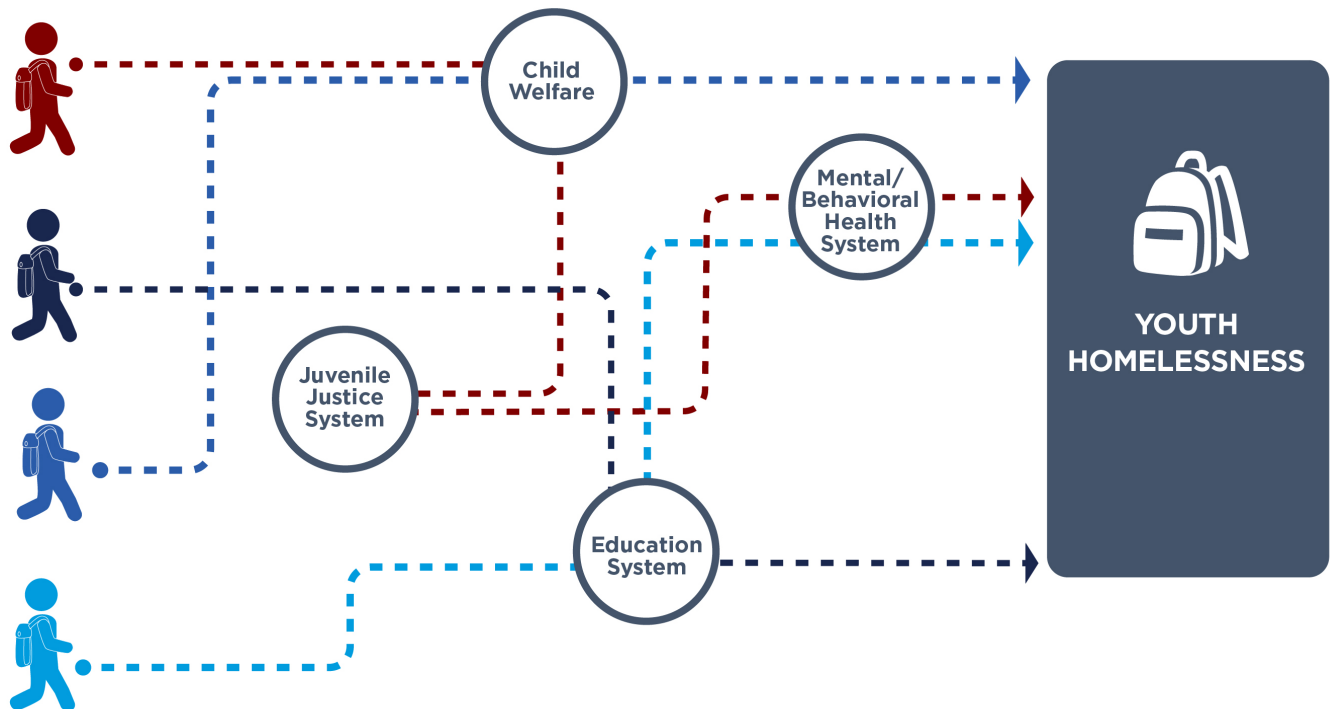
CATALOGUING DATA

The next step in data mapping is to catalogue relevant information about the data available from the data sources represented in the structural relationship map. This can be done by creating a data catalogue.

Data Catalogue

A data catalogue is a spreadsheet that captures high-level information about the data available from different data sources. Each row represents a different data source and each column represents a different piece of information. Figure 5 is a screen shot of a downloadable data catalogue template.

Figure 5. Example of a Data Catalogue

The data catalogue captures information about the data available from the data sources represented in the structural relationship map. It may be necessary to consult with the data source to obtain some of this information. Table 2 lists some of the questions that a data catalogue can answer.

Basic Information				Data pertaining to homelessness	History of the data
Agency Name	Database Name	Data Description	How are these data collected?	Are youth experiencing homelessness represented in these data?	First year for which data are available

Table 2. Key Questions Answered by Data Catalogue

Categories	Key Questions
Basic information	What agency owns these data? What is the name of the database? How and by whom are these data collected? Can the data be linked to data in other internal or external databases?
Homelessness	Do the data include indicators of current homelessness? Do these data include indicators of prior homelessness? How is homelessness defined?
Data history	For what years/months are these data available? How frequently are records updated? How frequently are records added?
Possible uses	Can these data be used to identify youth experiencing homelessness? Can these data be used to describe youth experiencing homelessness? Can these data be used to produce counts of homeless youth? Can these data be used to produce prevalence/incidence estimates? Can these data be used to explore pathways into homelessness?

Codebooks and Data Dictionaries

A codebook or data dictionary contains information about the content and layout of a dataset. This includes the name of each data element (that is, the field), a description of the data element, whether the data element is a unique identifier, the type of data (whether the data element is numeric or character), the length of the data element, and potential values. A codebook or data dictionary can help answer some of the questions listed in Table 2. Some codebooks or data dictionaries are publicly available; others may be provided by the relevant data source upon request. In some cases, a codebook or data dictionary may not exist and will need to be created. Table 3 is a template for a codebook or data dictionary.

Table 3. Codebook Template

Field	Description	Unique ID (Y/N)	Type	Length	Values

ACCESSING DATA

Once the potential data sources have been identified and their data have been catalogued, the next step is to request access to the data. Table 4 shows important questions to ask each data source about accessing the data.

Table 4. Key Questions Related to Accessing Data

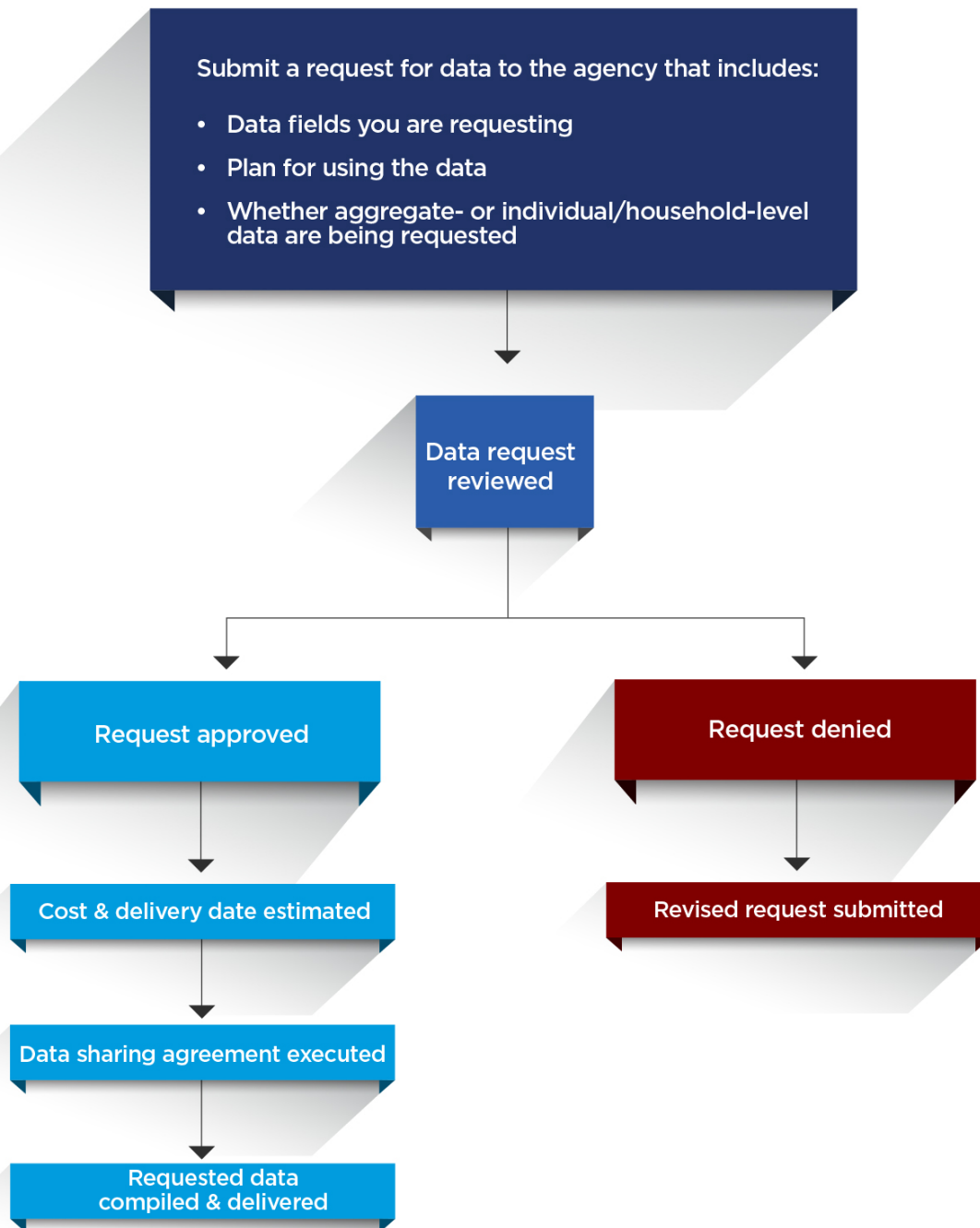| Key Questions |
|---------------|
| What data, including personal identifiers, can be shared? |
| What is the process for **requesting access to** these data? |
| What, if any, legal issues (for example, HIPPA, FERPA), must be considered when accessing these data?[19,20] |
| What is the **cost** associated with accessing these data? |
| How much **time** does it take to access these data? |

A data access flow chart illustrates the process for accessing data from a data source and can help answer the key questions listed in Table 4.

A data access flow chart is a visual representation of the process for requesting access to data from the data sources represented in the structural relationship map and the data catalogue. Figure 6 is a broad example of a data access flow chart illustrating the process of accessing administrative data from an identified data source. Different agencies and other organizations may have different processes and accessing data from some may be easier than accessing data from others.

Figure 6. Access Flow Chart Example

Submit a request for data to the agency that includes:

- Data fields you are requesting
- Plan for using the data
- Whether aggregate- or individual/household-level data are being requested

Data request reviewed

Request approved

Request denied

Cost & delivery date estimated

Revised request submitted

Data sharing agreement executed

Requested data compiled & delivered

Once the data have been requested and permission to use them has been obtained, users will likely need a **data sharing agreement** with the data source before the data can be shared.

Data Sharing Agreements

Data sharing agreements (DSAs) are formal contracts between two parties sharing data. DSAs typically specify the data that will be shared, the process that will be used to share the data, the purpose(s) for which the data can be used, the way the data will be kept secure, and the plan for data destruction. An example of a DSA template is linked here.

Secure Data Transfer

Secure file transfer protocol (SFTP) is a common method for securely transferring data. SFTP provides a high level of security because the data are encrypted while in transit. FileZilla is an open-source software that supports SFTP.

ASSESSING DATA QUALITY

The final step in data mapping is assessing the quality of the data. This involves carefully reviewing the data to identify any issues that need to be addressed before the data can be linked and analyzed. Table 5 lists questions to ask about the quality of the data.

**Table 5. Data Quality Questions**

Key Questions

- How much data are missing?
- What is the pattern of missing data (such as random or non-random)?
- Do any of the data fields include outliers?
- Do the data include duplicate records?
- Can any fields not be used due to concerns about data quality?

From Data Mapping to Data Linking and Analysis

One of the most important aspects of data mapping is that it provides a holistic understanding of the data prior to data linking and analysis. Once the quality of the data has been assessed, the next step is to determine how to combine the data from different sources to produce counts and descriptions of youth experiencing homelessness, estimate the prevalence and incidence of homelessness among youth, and explore pathways into homelessness from public systems. The process of combining data from different sources is referred to as data linking. Data linking requires a unique identifier that is common to the data sets being linked. Figure 7 shows an example of data from two sources being linked using a common unique identifier.

Figure 7. Linking HMIS Data to Child Welfare Data
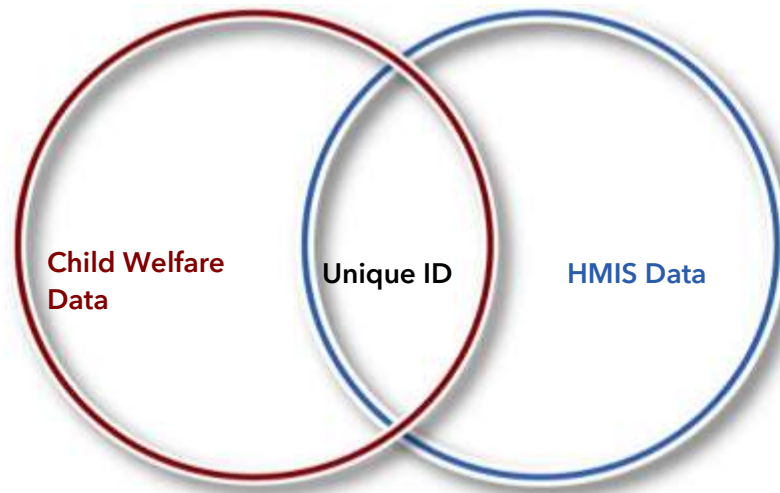
If the datasets being linked do not share a unique identifier, as is often the case, it may still be possible to link them using other identifying information that they have in common (for example, first and last name, birthdate, gender, race/ethnicity, and county of residence). One approach is deterministic record linkage. This approach compares a set of identifiers in one dataset to the same set of identifiers in another dataset; a link is made if they all agree. Another approach is probabilistic record linking. This approach uses multiple pieces of identifying information to calculate the probability that two records belong to the same individual.[21]

An important factor to consider when analyzing data from different sources is whether homelessness is consistently defined, and, if not, how to account for any inconsistencies in the analysis.

Conclusion

Analyzing linked administrative data can answer key questions about the number of youth experiencing homelessness, the characteristics of those youth, the prevalence and incidence of youth homelessness, and pathways into homelessness through public systems. However, analysis is just the last step in a process that includes figuring out which data to use, accessing those data, and assessing their quality. We developed this guide for communities that want to leverage their HMIS data paired with other administrative data to learn more about their homeless youth population. Communities can use the tools included in this guide to facilitate that process.

Statement of Independence and Integrity

Chapin Hall adheres to the values of science, meeting the highest standards of ethics, integrity, rigor, and objectivity in its research, analyses, and reporting. Learn more about the principles that drive our work in our [Statement of Independence](#).

Established in 1985, Chapin Hall partners with policymakers, practitioners, and philanthropists at the forefront of research and policy development by applying a unique blend of scientific research, real-world experience, and policy expertise to construct actionable information, practical tools, and, ultimately, positive change for children and families. Chapin Hall's areas of research include child welfare systems, community capacity to support children and families, and youth experiencing homelessness. For more information about Chapin Hall, visit www.chapinhall.org or @Chapin_Hall.

Acknowledgement and Disclaimer

Suggested Citation

Youngren, M., & Dworsky, A. (2024). *Using administrative data to understand youth homelessness: A data mapping guide*. Chapin Hall at the University of Chicago.

Correspondence

Madeline Youngren, Policy Analyst, myoungren@chapinhall.org

Amy Dworsky, Senior Research Fellow, adworsky@chapinhall.org

[1] Morton, M., Dworsky, A., Matjasko, J., Curry, S., Schlueter, D., Chávez, R., & Farrell, A. (2018). Prevalence and correlates of youth homelessness in the United States. *Journal of Adolescent Health, 62*, 14–21. doi: 10.1016/j.jadohealth.2017.10.006

[2] Dworsky, A., Napolitano, L., & Courtney, M. 2013. Homelessness during the transition from foster care to adulthood. *American Journal of Public Health, 103*, 318–323.

[3] Dworsky, A., Gitlow, E., Horwitz, B., & Samuels, G. (2019). *Missed opportunities: Pathways from foster care to youth homelessness in America*. Chapin Hall at the University of Chicago.

[4] Feng, H., Harty, J., Okpych, N., & Courtney, M. (2020). *Predictors of homelessness at age 21*. Chapin Hall at the University of Chicago.

[5] Fowler, P., Toro, P., & Miles, B. W. (2009). Pathways to and from homelessness and associated psychosocial outcomes among adolescents leaving the foster care system. *American Journal of Public Health*, *99*, 1453–1458.

[6] Hernandez, H., Danielson, T., Mayfield, J., Black, C., & Felver, B. (2023). *Homelessness among youth exiting systems of care in Washington State*. Washington State Department of Social and Health Services, Research Data Analysis Division.

[7] Morton, M., Dworsky, A., & Samuels, G. M. (2017). *Missed opportunities: Youth homelessness in America. National estimates*. Chapin Hall at the University of Chicago.

[8] Narendorf, S., Brydon, D., Santa Maria, D., Bender, K., Ferguson, K., Hsu, H., Barman-Adhikari, A., Shelton, J., & Petering, R. (2020). System involvement among young adults experiencing homelessness: Characteristics of four system-involved subgroups and relationship to risk outcomes. *Children and Youth Services Review, 108,* 1–9.

[9] Whitbeck, L., Lazoritz, M., Crawford, D., & Hautala, D. (2016). *Street Outreach Program Data Collection Study Final Report*. U.S. Department of Health and Human Services, Administration for Children and Families, Administration on Children, Youth, and Families, Family and Youth Services Bureau.

[10] Yoder, J., Bender, K., Thompson, S., Ferguson, K., & Haffejee, B. (2014). Explaining homeless youths' criminal justice interactions: Childhood trauma or surviving life on the streets? *Community Mental Health Journal, 50*, 135–144.

[11] Edidin, J., Ganim, Z., Hunter, S., & Karnick, N. (2012). The mental and physical health of homeless youth: A literature review. *Child Psychiatry and Human Development, 43*, 354–375.

[12] Hodgson, K., Shelton, K., van den Bree, M., & Los, F. (2013). Psychopathology in young people experiencing homelessness: A systematic review. *American Journal of Public Health, 103*, 24–37.

[13] Perlman, S., Willard, J., Herbers, J., Cutuli, J., & Eyrich Garg, K. (2014). Youth homelessness: Prevalence and mental health correlates. *Journal of Social Work Research, 5*, 361–377.

[14] Auerswald, C., Lin, J., & Parriott, A. (2016). Six-year mortality in a street-recruited cohort of homeless youth in San Francisco, California. *PeerJ, 4,* e1909.

[15] Dworsky, A., Morton, M., & Samuels, G. (2018). *Missed opportunities: Pregnant and parenting youth experiencing homelessness in America*. Chapin Hall at the University of Chicago.

[16] Edidin, J., Ganim, Z., Hunter, S., & Karnick, N. (2012). The mental and physical health of homeless youth: A literature review. *Child Psychiatry and Human Development,* 43, 354–375.

[17] Medlow, S., Klineberg, E., & Steinbeck, K. (2014). The health diagnoses of homeless adolescents: A systematic review of the literature. *Journal of Adolescence, 37*, 531–542.

[18] Kulik, D., Gaetz, S., Crowe, C., & Ford-Jones, E. L. (2011). Homeless youth's overwhelming health burden: A review of the literature. *Paediatrics & Child Health*, *16*, e43–e47.

[19] The Health Insurance Portability and Accountability Act of 1996 (HIPAA) is a federal law that requires the creation of national standards to protect sensitive patient health information from being disclosed without the patient's consent or knowledge. See Centers for Disease Control and Prevention. (2022). *Health Insurance Portability and Accountability Act of 1996 (HIPAA)*. Centers for Disease Control and Prevention. https://www.cdc.gov/phlp/publications/topic/hipaa.html

[20] The Family Educational Rights and Privacy Act (FERPA; 20 U.S.C. § 1232g; 34 CFR Part 99) is a federal law that protects the privacy of student education records. The law applies to all schools that receive funds under an applicable program of the U.S. Department of Education. See U.S. Department of Education. (2021). Family Educational Rights and Privacy Act (FERPA). https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

[21] For more information about these two approaches to record linkage, see Goerge, R. M., & Lee, B. J. (2002). Matching and cleaning administrative data. In M. Ver Ploeg, R. A. Moffitt, C. F. Citro (Eds.), *Studies of welfare populations: Data collection and research issues* (pp. 197–219). https://nap.nationalacademies.org/read/10206/chapter/9.