EXPANDING MOBILITY:
The Power of Linked Administrative Data + Integrated Data Systems

AISP

Acknowledgments

"Expanding Mobility: The Power of Linked Administrative Data and Integrated Data Systems" was created by Actionable Intelligence for Social Policy (AISP) at the University of Pennsylvania. The report is the first in a series on administrative data use for economic mobility research and analysis. It was developed based on findings from our 2020 national survey and landscape analysis of data integration efforts across the US as well as from more than a decade of working with and learning alongside our Network and Learning Community Sites.

AISP is supported by grant funding from the Bill & Melinda Gates Foundation and the Annie E. Casey Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the funders.

Many thanks to everyone who made it possible.

AISP Contributors
Lead authors:
Della Jenkins and
Emily Berkowitz

TC Burnett
Dennis Culhane, PhD
Amy Hawn Nelson, PhD
Matthew Katz
Kristen Smith
Sharon Zanti

Suggested Citation: Jenkins, D., Berkowitz, E., Burnett, TC, Culhane, D., Hawn Nelson, A., Katz, M., Smith, K., & Zanti, S. (2021). Expanding Mobility: The Power of Linked Administrative Data and Integrated Data Systems. Actionable Intelligence for Social Policy. University of Pennsylvania.

Introduction

For too long, conceptions of economic mobility have been limited to individual outcomes in the areas of educational achievement and earnings. While there are exceptions, this emphasis on the value of degrees and income often fails to capture the range of individual experiences beyond school and work that also matter to "success." At the same time, government policies have overemphasized the ability of individuals and households to "pull themselves up by their bootstraps," obscuring the major role structural forces and market dynamics play in determining individual outcomes. Doing so also overlooks how a person's year and place of birth—as well as that of their parents and ancestors—has a sizeable impact on their outcomes in adulthood. Opportunity (or lack thereof) compounds, concentrating power and choice among those born with resources. A person's well-being and ability to achieve upward mobility depend on their whole ecosystem: their relationships, their sense of community, and more materially, their access to basic needs (e.g., transportation, nutritious food) and assets (e.g., stable housing, credit).

Data can be employed to advance common narratives that characterize poverty as an "individual deficit" and justify long-standing individualistic approaches to poverty. Therefore, how metrics are measured and contextualized matters as much as the metrics themselves. In 2018, the U.S. Partnership on Mobility from Poverty developed an expansive definition of mobility from poverty that takes major steps beyond individual economic and financial well-being to consider power and autonomy, as well as community-level factors. In 2019, the Urban Institute established a cross-disciplinary group of academic experts to develop a framework for understanding and promoting mobility out of poverty. Their resulting report—"Boosting Upward Mobility"—includes mobility metrics that measure conditions that affect long-term community outcomes, rather than short-term individual outcomes. Further, the framework considers outcomes that can be influenced by state and local government policy changes.

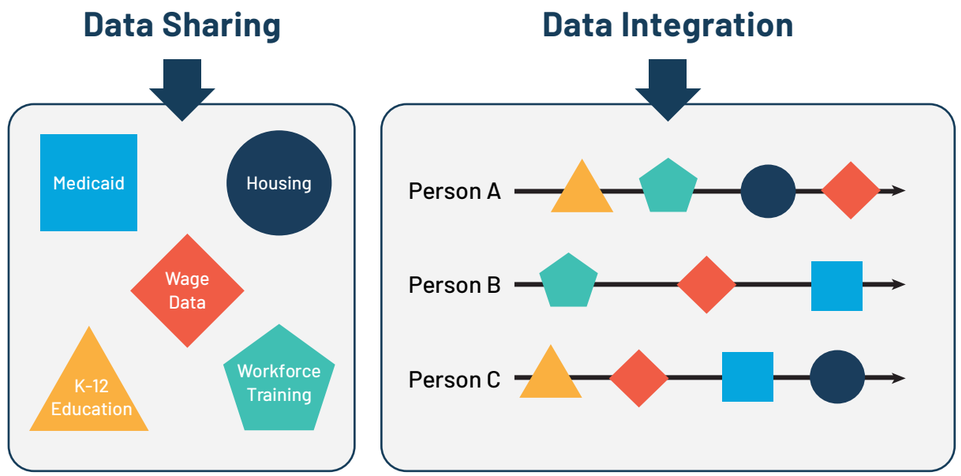
> ***What are administrative data?***
> **Data that are collected during the routine process of administering public programs, but that can also be repurposed to support evaluation, analysis, and research.**

Integrated administrative data are essential to power a practice shift toward measuring long-term, contextual, actionable metrics like those identified in "Boosting Upward Mobility." Linking administrative data across sectors provides at once a more granular and expansive picture of what drives or hinders mobility out of poverty. Individuals engaged in public programs can be viewed across the lifespan to better understand how access to resources and opportunity helps foster different outcomes. Moreover, these findings on individual trajectories can be analyzed to better understand service utilization patterns so that underserved groups and people with more complex needs can be identified and better served. Such findings may also inform policy shifts that put public dollars to more efficient use and generate better downstream outcomes, resulting in material improvement for individuals and communities.

This brief describes how linking administrative data can expand traditional measures of mobility for research and action, provides examples of the types of economic mobility research questions that are only answerable using linked administrative data, and describes how analysis can be deepened using spatial and multi-generational perspectives. In addition, we discuss how the field of economic mobility research benefits when state and local governments are resourced to build systems that enable routine reuse of linked data. Finally, we end with a summary of the opportunities that exist to build on data capacity already developed by state and local governments across the US to better understand the policies that support pathways out of poverty. Now more than ever, governments, research partners, and stakeholders can come together to make use of the data already collected by social service programs to generate evidence-based approaches to expanding mobility.

Benefits of Linked Administrative Data Use For Economic Mobility

Administrative data can be shared or integrated in order to elicit insights on economic mobility. Sharing and integrating data are related but distinct. When data are shared, they can be used to power dashboards and social indicators projects that help stakeholders better understand mobility at the community level. When data are integrated or linked at the individual level across sectors, they reveal trends, patterns, and insights about mobility that would otherwise go undetected.



Data sharing is the practice of providing partners with access to information (in this case, administrative data) that they cannot access in their own data systems. Data sharing allows stakeholders to learn from each other and collaborate on shared priorities.

Data integration is a more complex type of data sharing that involves record linkage, which refers to the joining or merging of data based on common data fields. These data fields can include personal identifiers, such as name, birth date, social security number, or a common encrypted "unique ID" that is used to link or join records at the individual level. Personal identifiers may or may not be removed, depending on the use case and legal framework.

Learn more in AISP's Introduction to Data Sharing & Integration.

In this brief, we focus on the value-add of integration, or linkage. With linked individual-level administrative data, mobility researchers can:

- Observe trends and mobility trajectories across a population
- Better identify and understand the service use patterns of population subgroups
- Measure and evaluate specific program impact on mobility over time
- Evaluate the impact of broader social policies on well-being and mobility more holistically

What's more, researchers can deepen any of the four types of analysis outlined above by conducting an analysis that is **spatial** or **multi-generational.**

Below, we take a deeper dive into these applications and methods. We explore how integrated data can be leveraged for economic mobility research in order to reveal increasingly detailed and actionable information about risk, resiliency, and outcomes, using the example of a Philadelphia high school cohort.

Applications of Linked Administrative Data for Economic Mobility Research

1. Start big: Observe trends and mobility trajectories across a population

Integrated administrative data offer an expansive view of individuals and their ecosystem when used for observational, population-based studies. Because individual-level data are aggregated across services and time, researchers and analysts can use latent growth curve modeling to understand the mobility trajectories within a population. Whether that population is as large as a state or as small as a public high school graduating class, incorporating information from multiple services and datasets can broaden our understanding of the characteristics associated with various trajectories. This type of descriptive analysis can be useful in understanding the pathways of mobility. It can also serve as an important baseline on which to build future research on subpopulations and against which to measure the efficacy of future interventions.

EXAMPLE

Looking at all Philadelphia high school graduates and attendees from 2008 to 2010, where are they by age 25 in regard to workforce participation and earnings? What trajectories are associated with different types of educational attainment or workforce training? How do these trajectories differ by high school? By gender? By race?

POSSIBLE DATA ELEMENTS USED IN LINKAGE:

Vital Statistics

Birth records

Education

Early childhood education, Lead exposure, K-12 attendance and achievement, Student learning disability indicators, Child welfare involvement, FAFSA, Standardized test scores, Postsecondary education, Workforce training, Student disability status

Economic Security

NDNH, Unemployment insurance/wage date, SSI, Earned Income Tax Credit, Workforce training programs, Special Supplemental Nutrition for Women Infants & Children, Temporary Assistance for Needy Families

2. Drill down: Better identify and understand the service use patterns of population subgroups

Within any population, there are individuals who experience barriers, challenges, or vulnerabilities related to achievement and well-being. Linked administrative data enable analysis that segments populations in order to better understand need, service gaps, and multi-system service utilization. Population subgroups in economic mobility research commonly include young adults with histories of service involvement (i.e., social services, foster care, mental health services), people with disabilities, students with high rates of absenteeism, and those exiting detention or incarceration, among others. By understanding the many touch points that individuals in vulnerable subgroups may (or may not) have with agencies and programs, governments can improve service coordination, quality, and efficiency. What's more, researchers can better understand people's layered experiences and assess drivers of both risk and resiliency.

EXAMPLE

Within the population of Philadelphia high school graduates and attendees, can we identify vulnerable populations (i.e. students receiving special education services, students experiencing homelessness, pregnant and parenting students)? What services are they accessing? In which combinations? What are their mobility trajectories? What are their outcomes compared to the general population?

POSSIBLE DATA ELEMENTS USED IN LINKAGE:

- **Vital Statistics**
- **Education**
- **Economic Security**
- **Homelessness/Housing**

  Homelessness Management Information System (HMIS), Public Housing Authority, Education records indicating homelessness (McKinney-Vento), State housing subsidy

NOTE: *Linked administrative data also provide a method for targeting services or precision interventions to subgroups. In order to go one step further and utilize linked data to make direct contact with individuals in subgroups, personal identifiers are needed. The use of identifiable integrated data requires different legal approvals and technical approaches than when sharing and linked data without identifiers. As a result, the use of administrative data that includes personally identifiable information is a distinct use case and should only be undertaken by mature data integration efforts.*

3. Look ahead: Measure and evaluate specific program impacts on mobility over time

It's a classic dilemma: the outcomes we care about most—those that capture whether a program had the intended long-term impacts on mobility—are often the hardest to measure. Students graduate, program participation ends, and we lose track of where people end up. By linking administrative data across agencies, governments and researchers get a much clearer view of what services or programs people participate in and how they fare over time. This can be accomplished solely with administrative data sources, or it may involve matching survey data with administrative records to find out what happened to participants in a cohort long after original data collection for a study is complete.

EXAMPLE

Following up with the cohort of Philadelphia high school graduates and attendees at age 30, where are they now in terms of higher education attainment, workforce participation, and earnings? Did those who enrolled in workforce training programs see improved earnings? Job stability? How do the outcomes of those who received 2-year degrees compare to those who received 4-year degrees? By institution? For those who were incarcerated at age 21 and received re-entry services focused on job placement, where are they now?

POSSIBLE DATA ELEMENTS USED IN LINKAGE:

- **Vital Statistics**
- **Education**
- **Economic Security**
- **Homelessness/Housing**
- **Legal System and Law Enforcement**

  Adult court, Adult probation, Local jail, State Corrections, Law enforcement (i.e., arrests, citations, incidents)

4. Broaden the lens: Evaluate the impact of broader social policies on well-being and mobility more holistically

Expanded definitions of mobility are not limited to indicators of educational attainment and earnings. To implement an expanded definition, data models should consider health and well-being more broadly, as well as the social policies that may affect them. For example, policies like Medicaid expansion, raising the age of criminal involvement in the adult system from 16 to 18, or extending the age of emancipation from foster care from 18 to 21 are not "economic" per se, yet they may have significant impacts on health, well-being, workforce engagement, and long-term success. Constructing comparison groups using cross-sector and longitudinal data more readily enables evaluation of social policy changes across a broad array of outcomes.

EXAMPLE

How do Philadelphia high school graduates look at age 30 in terms of broader metrics of well-being and mobility (physical and mental health, housing stability)? How have social policies impacted their trajectories? Of those who experienced housing instability, did those who received rent subsidies maintain more stable employment over time? Did they experience lower rates of substance misuse? Of those who were incarcerated at some point, did those who received record expungement see more stable employment over time? Better long-term health outcomes?

POSSIBLE DATA ELEMENTS USED IN LINKAGE:

- **Vital Statistics**
- **Education**
- **Economic Security**
- **Homelessness/Housing**
- **Legal System and Law Enforcement**
- **Health**

  Medicaid, All payer health claims, Community health centers, Subtance use, Behavioral and mental health, Emergency Medical Services

Spatial and Multi-Gen Analysis of Mobility

In addition to helping us start big, drill down, look ahead, and broaden the lens, linked administrative data can help us look at mobility spatially or multi-generationally.

1. Spatial Analysis

There is broad consensus that upward mobility requires access to opportunities and resources at an individual and neighborhood level. This is because inequality is organized, maintained, and reinforced spatially. In order to deepen our analysis of mobility, it is important to leverage both "people data" and "place data." When linked administrative data on people is combined with data on places and spaces, researchers can study the historical, built environment, and social environments that impact mobility over time.

Using individual-level data may make findings more nuanced and actionable. For more, see AISP's report, "Expanding Mobility: The Power of Linked Administrative Data for Spatial Analysis."

Example:

> Do Philadelphia high school graduate mobility trajectories vary as a function of neighborhoods? Which neighborhood characteristics influence trajectories? Physical conditions such as housing quality and affordability? Access to resources like public libraries or recreation centers? Exposure to lead or environmental toxins? Or peer influences, such as rates of chronic absenteeism and legal system involvement?

2. Multigenerational Analysis

Research consistently confirms that children born into poor families are more likely to be poor as adults than their peers whose families have greater financial resources. To put it simply, money begets opportunity and resources. The resulting "path dependency" of socioeconomic status is further compounded by structural barriers to mobility within one's community and immediate environment, such as school and neighborhood segregation. There is an existing literature that relies upon the use of linked administrative data to describe individual trajectories using a "life course" analysis. Multigenerational (multi-gen) linkage enables us to take this work one step further, capturing intergenerational impacts on multiple people over time. Importantly, multi-gen analysis using linked administrative data raises complex questions about how we define family or household—the answer to which inevitably changes depending on the research question and data available to answer it. By exploring these questions and putting emerging methods into action, we can deepen our understanding of human development, well-being, and mobility. For more, see AISP's report, "Expanding Mobility: the Power of Linked Administrative Data for Multi-Gen Analysis."

Example:

> Do Philadelphia high school graduate trajectories vary as a function of intergenerational influences, such as parental employment patterns, or household composition (single versus two parent households)? Does parental incarceration impact high school graduation? Is a person more likely to be incarcerated as an adult if one of their parents was incarcerated during their childhood? Does the presence of grandparents in a single parent household improve mobility outcomes?

Leveraging Linked Administrative Data For Economic Mobility

Key Considerations for Data Access and Use

Of course, accessing, linking, and using administrative data to surface greater understanding of economic mobility takes time, effort, and expertise. If you are a policymaker, researcher, or other stakeholder interested in administrative data reuse for this purpose, here are a few key considerations before getting started.

1. Administrative data are collected for operations, not research and analysis.

"Administrative data" refers to the information collected during the course of routine operations. While these data can be repurposed for research and analysis, doing so requires carefully assessing whether the quality of the data is sufficient for the intended use. This step is key, as unaddressed data quality issues will impact the reliability and validity of results. Administrative data can also be misinterpreted if care isn't first taken to understand the context in which they were collected. Such misinterpretations could include an inappropriate analytic plan (e.g., use of predictive tools with data of poor quality), misuse of a variable (e.g., incorrect assumption that PRGENT refers to the date of program entry, rather than the date of program entrance exam), or the inclusion of incorrect assumptions when explaining outcomes (e.g., explaining a reduction in out-of-school suspensions as being a positive indicator of climate, without knowing of a code of conduct revision that changed reporting of suspensions). Collaborating with agency staff who have both analytic and programmatic expertise—those who "know" the data best—can help bridge key gaps in understanding and avoid misinterpretation issues.

2. Datasets only show where the light is shining.

Administrative data only contain information on individuals served by public programs and agencies, which automatically limits who is captured. While birth, Department of Motor Vehicles, and tax records are some of the most complete datasets for studying a population-level cohort, since they capture a wide range of individuals regardless of socioeconomic status or contact with public programs, they are not without limitations—for example, tax records often do not include people who are insecurely housed or homeless, and birth records do not capture immigrants. Further, these datasets tend to be more difficult to access for research purposes, and, as a result, are less frequently used by data integration efforts in the US. Researchers, analysts, and other stakeholders using administrative data to answer research and policy questions must be thoughtful and strategic in considering the associated risks or limitations of a given dataset.

3. Multiple methods and types of research are needed to capture a holistic definition of mobility.

Questions about mobility are often highly subjective and, without appropriate context, can be evaluated in ways that reinforce bias or stereotypes. For example, as a result of the legacy of racist policies in the US, including redlining and government-sanctioned housing segregation, Black homeownership rates are lower than White homeownership rates. Without the historical context, a project that considers the role of homeownership or family assets in economic mobility may interpret this as a function of individual mobility factors, rather than the historical factors that inhibit homeownership and wealth accumulation. Analyses could include this context by examining whether people's neighborhoods belong to historically redlined areas, or whether their parents come from areas that were historically redlined.

Strategic partnerships with private data vendors, who can track residential addresses historically, sometimes help to fill gaps in dimensions of economic mobility that are not available in more recent government administrative data. Administrative data can also be supplemented and contextualized using other data sources, such as surveys and qualitative research, in order to ensure that lived experience is reflected. Learn more about how to involve community in your data integration efforts in AISP's A Toolkit for Centering Racial Equity Throughout Data Integration.

Housing First study in North Carolina uses mixed methods to evaluate outcomes

Housing First Charlotte-Mecklenburg (HFCM) is a multi-sector public-private initiative focused on ending chronic homelessness through wider implementation of the Housing First permanent supportive housing model. Since 2014, HFCM has housed over 1,000 people experiencing chronic homelessness. A 2020 evaluation sought to better understand the outcomes associated with HFCM program involvement using quantitative and qualitative methods.

Specifically, structured interviews, observations, and document reviews were conducted by evaluators. The Institute for Social Capital (ISC), the IDS serving Mecklenburg County, conducted the administrative data linkage and quantitative analysis using linked data across 14 agencies. Linked data enabled HFCM to evaluate social service utilization patterns for their participants before and after housing placements, providing otherwise hidden insights on the totality of experiences associated with chronic housing challenges and, conversely, the benefits supportive permanent housing provides clients. A mixed-methods approach also allowed evaluators to uncover inequities related to race, ethnicity, age, and gender as they relate to HFCM's identification of individuals with chronic housing needs and the program's service allocation. These findings allow HFCM to more equitably structure permanent supportive housing services and target them to individuals with housing needs throughout the county.

4. Accessing and using wage data can be challenging.

State and federal wage data are critical to tracking employment and earnings over time, yet they are among the more challenging to access. The federal tax system and the state Unemployment Insurance (UI) system are the two primary sources of earnings data in the US, and there are a few potential avenues to access them (see Mathematica's Data on Earnings report). In some states, UI data are commonly used to assess education and training program outcomes, and procedures are in place for matching these records. Additionally, the US Department of Labor has issued guidelines that explain the multiple ways in which education and wage data can be linked and the various legal auspices under which these data can be shared with analysts. However, capacity constraints and perceived legal barriers often limit wage record linkage opportunities in other domains. IDS, which are explored in detail in the following section, represent an approach whereby state wage record custodians can maintain authority over their data while also supporting the secure linkage of wage records for high-impact economic mobility research.

5. Many datasets lack information on both children and caregivers.

Organizing and linking administrative data into households and families in order to enable multi-generational analysis is complicated, and datasets that facilitate this linkage are limited. The terms "household" and "family" themselves are fraught. Datasets that capture information on "household roster" (e.g., Temporary Assistance for Needy Families, Supplemental Nutrition Assistance Program (SNAP), Supplemental Nutrition Program for Women, Infants and Children, Department of Housing and Urban Development) are most feasible and frequently used for multi-generation linkage. However, these datasets do not share common definitions of household, families, or benefits groups, and all are limited in that they capture only who resides together at a given point in time. Similarly, other datasets that include address data (such as school records) can be matched to construct family units based on who shares a residence at a given point in time, but there is the risk that there are multiple units at the same address, or families "doubling up" (multiple households living under the same roof). Importantly, this is an area of rapid innovation, and many stakeholders are working to document options and develop new methods for linkage across households/families and generations. To learn more, see the AISP report, "Expanding Mobility: The Power of Linked Administrative Data for Multi-Gen Analysis."

Benefits of an Integrated Data System: Moving From Ad Hoc to Routine Use

To best meet the needs of households and communities experiencing poverty, insights about mobility must be actionable and data use must be routine. Unfortunately, ad hoc research projects using linked data are often conducted without any mechanisms in place to alter programs or policies in response to findings. What's more, when projects are conducted on a one-off basis, the legal agreements and linkage strategies must be renegotiated each time, severely limiting the frequency of data use and causing additional administrative burden. However, by investing in the relational and technical capacity to use linked administrative data routinely, governments and university partnerships

can substantially reduce the time and cost associated with these transactions. At AISP, efforts that build this routine capacity are referred to as integrated data systems (IDS), or, more broadly, data integration efforts. They may also be known as data collaboratives, State Longitudinal Data Systems, or data hubs. Defining features of such efforts include engagement across data contributing agencies, streamlined legal frameworks, and strong protocols to govern how data are used and protected. For more on best practices and actionable strategies to advance data sharing, see AISP's Quality Framework.

By decreasing barriers to data access, integration, and use, IDS also allow policies and programs aimed at increasing economic mobility to be better implemented, evaluated, and improved upon—what may be referred to as the cycle of continuous improvement or a "knowledge to practice" cycle. Over time, evidence can even be used to anticipate service needs in order to proactively design programs to address identified gaps. Finally, systematizing data use also generally contributes to more efficient and evidence-based governing, which has proven to benefit both leadership and constituents (see Results for America's Blueprint for Delivering Results in State Government).

Conclusion

Linked administrative data and integrated data systems are an essential tool for nuanced and comprehensive economic mobility research. Linked data allow us to better capture the complex experiences of people over time and across domains so that we can craft equitable solutions at the systems level. While the work of sharing and linking data on economic mobility can be challenging, there are many existing data resources to leverage and expand toward this goal. Findings from a national survey of data integration efforts, conducted by AISP in February 2020, show that there has been a sizeable increase in the field of administrative data reuse; of the 63 sites surveyed, nearly half (46%) launched their effort in the past five years.

Results from the survey, which captured information on efforts to make ethical, efficient data use more routine, demonstrate a clear opportunity to build on the growing data capacity of state and local efforts for mobility research and programs. Of particular interest in this context, 40% of sites surveyed reported access to some data on economic success (e.g., wage data, supplemental security income, SNAP), which is necessary for any integration work focused on mobility—from basic descriptive reports to more in-depth analysis on what helps or hinders mobility. Survey analysis also found that the most commonly accessed data categories were (in order) early childhood, education, health data, and homelessness and housing, suggesting sufficient capacity for cross-sector linkage that provides a holistic view of mobility. Such efforts, if properly resourced and sustained, have the potential to scale for impact and truly reshape the field of economic mobility research.

THIS REPORT is the first in AISP's 3-part "Expanding Mobility" series, which explores how linked administrative data can be used to deepen understanding of economic mobility. The two complementary reports, linked below, focus on the use of integrated administrative data for mobility research and multi-gen analysis.

- *Expanding Mobility: The Power of Linked Administrative Data for Spatial Analysis*
- *Expanding Mobility: The Power of Linked Administrative Data for Multi-Gen Analysis*

To learn more about building and scaling IDS, visit the AISP website and view our Quality Framework.


AISP