

CROSS-STATE DATA SHARING: POTENTIAL BENEFITS AND SURMOUNTABLE BARRIERS

MAY 2021



D4
AD

 **NCHEMS**

ABOUT THIS BRIEF

Cross-state Data Sharing: Potential Benefits and Surmountable Barriers explores important considerations for entering and maintaining strong data sharing partnerships. This brief offers best practices and advice for organizations in creating the agreements, navigating regulation, and developing governance structures.

The author, Patrick Lane, is vice president for policy analysis and research at the Western Interstate Commission for Higher Education (WICHE), an organization that has worked to facilitate cross-state data sharing of education and employment data in addition to other efforts to improve access to and success in postsecondary education. Information contained in this brief is drawn from lessons learned during the Multistate Longitudinal Data Exchange Effort as well as general research. The views presented in this brief are those of the author and do not represent Schmidt Futures, Lumina Foundation, Walmart Foundation, or the Walton Family Foundation.

ABOUT DATA FOR THE AMERICAN DREAM

Data for the American Dream (D4AD) is a consortium bringing together Schmidt Futures, Lumina Foundation, Walmart Foundation, and the Walton Family Foundation. D4AD currently funds pilot initiatives in three states (Colorado, Michigan, and New Jersey) that will help provide low-income, lower-skilled, underemployed, and unemployed workers access current and actionable data, enlisting local case managers from public and private agencies to counsel job seekers, help them access needed services, and reach the most underserved populations.

The National Center for Higher Education Management Systems (NCHEMS) is the implementation partner of D4AD. NCHEMS is a private nonprofit organization whose mission is to improve strategic decision making in postsecondary education for states, systems, institutions, and workforce development organizations in the United States and abroad.





BACKGROUND AND INTRODUCTION

Data for the American Dream is an ambitious initiative to take advantage of information about education and training opportunities, employment outcomes, and labor markets to give low-income, unemployed, and underemployed individuals better career opportunities. The theory behind this effort is that these individuals will be better able to make life-changing decisions about which education and training opportunities to pursue or what career pathway to follow if they have better data and information about job openings available and the outcomes of related education and training programs.

This work is partnering with public agencies and private organizations to solve significant challenges around data infrastructure and governance, outreach and engagement, and behavioral economics. The work is funded by Schmidt Futures, Lumina Foundation, Walmart Foundation, and the Walton Family Foundation, with the National Center for Higher Education Management Systems serving as the implementation partner.

The effort has selected three initial grantees working in Colorado, Michigan, and New Jersey. One commonality among the three efforts is an interest in including employment data from other states to fill data gaps in their own systems. Adding these data is important but doing so entails significant barriers and challenges. The benefits of including such data will be greater in locations where there is significant cross-state migration and multi-state employment markets. Efforts to obtain such data are important, but also could potentially distract attention and bandwidth from the broader efforts underway to better serve the populations of interest. The degree to which that may be true is likely related to the following:

- The effort required to obtain the data from other states and effectively combine it with existing state data.
- The extent to which regional economic activity lies beyond the borders of the state of interest.
- The relative share of graduates who find employment in another state, especially if that varies substantially by program (which is, of course, not something that is generally well known without the sharing of data across state lines).

This document provides background information on issues identified by other efforts to share protected administrative data across state lines and ways those issues have been addressed (with varying levels of success).

Beyond administrative data – particularly state Unemployment Insurance records that are used to provide salary outcomes for different credentials – there are certainly other opportunities and data sources that may be useful for states and others engaged in this work.



BENEFITS

Cross-state data sharing is undeniably an important piece of providing accurate information about the labor-market outcomes associated with postsecondary credentials. Incorporating data from other states not only improves the accuracy of consumer information available to individuals deciding between different options, but it also provides better metrics for state policymakers and others to judge the effectiveness of programs.

Most states are currently able to match wage data from their Unemployment Insurance data systems with student records to determine earnings estimates for different postsecondary credentials.¹ If workers cross into another state for employment, the original state does not have wage data in its own systems, which can lead to some uncertainty about the true employment outcomes associated with different credentials. This can affect the availability of information useful to prospective students about the likely outcomes of an education or training program. It also can limit the information about program quality that may be available to state agencies that authorize and oversee education and training providers.

A pilot project carried out by the Western Interstate Commission for Higher Education (WICHE) shows that there can be meaningful differences in earnings for graduates of certain programs depending on whether they work in states where they earned their degree or cross state lines for employment.² Similarly, partnerships between the U.S. Census Bureau and several states have shown differences in earnings of degree completers when using national data compared to a single state's data.³ Including the data from other states tends to increase the earnings connected to different degrees, as mobility and higher earnings are correlated.⁴

A key question is the degree to which the information that grantees will provide through their tools may be incomplete and potentially misleading without including data from other states. In theory, without these data, the information provided to the population of interest by the grantees could underestimate the earnings associated with certain careers. If certain industries or major employers in a regional economy happen to be concentrated on one side of the state line, then the aligned training providers on the other are also likely to see artificially low employment rates of their graduates, in addition to this effect on wages. Filling these gaps in the data caused by migration and cross-state commuting could therefore provide potential students with more accurate information to use in making a choice among programs. It may do so either by providing a better, more accurate indicator of the placement rates of program graduates and of the wages a program graduate might reasonably expect. These could expand the range of career pathways a prospective student or jobseeker might consider—bringing a larger number of jobs and related training opportunities within relatively easy reach in a regional economy that spans multiple states.



Similarly, these data can help state agencies better evaluate the quality and relevance of education and training programs they authorize and/or fund. If, as an example, a training program leads to high-paying jobs, but many of those are across state lines, evaluations of the program without cross-state data could show weak employment numbers and limited earnings benefits, leading policymakers to question its efficacy and potentially shift funding and other resources to other programs. Certainly states can decide that in-state outcomes are more important than out-of-state employment, but currently they do not have the necessary information to understand and examine these nuances. But, if the primary goal of training and education programs is to secure stable and meaningful employment for participants, then securing any employment should be seen as a positive program outcome, even if there are some benefits to neighboring states.

It is impossible to answer exactly how much states would benefit from adding other states' data, but a closer examination of migration and commuting data can help shed some light on the issue.

Migration. Moving from one state to another varies significantly by state. Tax data from the Internal Revenue Service show 4.1 percent of Colorado taxpayers moved to a new state in 2018, while 2.7 percent of New Jersey residents and 1.7 percent of Michigan residents migrated.⁵

While these numbers may seem relatively small, it is important to recognize that the impact on data quality can have a long-term and cumulative effect. Such churn can severely impair longitudinal data analyses of employment outcomes related to education and training opportunities. Data from Maryland show how a cohort of community college graduates gradually disappeared from the state's Unemployment Insurance data file over time, leaving approximately half of the original cohort with wage records in Maryland after 25 years.⁶

It is worth noting that the population of interest in Data for the American Dream may be less likely to migrate, as research shows that the rate at which individuals migrate to another state increases with education levels.⁷

Commuting. Cross-state migration is only one potential cause of data gaps. In locations near state borders, individuals may live and seek training in one state, but find employment in a nearby state. This is especially likely when an economic region includes more than one state, such as when an urban area straddles a multistate border.

This cross-state commuting varies significantly by types of jobs sought and proximity to other states and labor markets, as one would expect. Table I below shows percentage of a state's residents working in another state. New Jersey, for example, shows the highest percentage of residents commuting to other states at 18 percent.⁸ For positions earning \$40,000 or less annually, that number drops to 11 percent.⁹ Areas closest to state borders show slightly higher rates of commuting, but the amount of this increase varies by state. The four New Jersey counties closest to New York, for example, have 20 percent of their residents crossing state lines for employment. Southwestern Michigan counties covered by the Michigan Works! Southwest workforce area have 3.5 percent of their residents commuting to another state.¹¹

| TABLE I: COMMUTING RATES BY STATE | | | | |
|-----------------------------------|----------------------------------|------------------------------|----------------------------------|------------------------------|
| State | All Jobs | | Jobs Earning < \$40,000 | |
| | Living and employed in the state | Commuting across state lines | Living and employed in the state | Commuting across state lines |
| Colorado | 98.3% | 1.7% | 98.1% | 1.9% |
| Michigan | 96.7% | 3.3% | 96.6% | 3.4% |
| New Jersey | 81.9% | 18.1% | 88.9% | 11.1% |

Source: U.S. Census Bureau¹²

As Table I shows, commuting across state lines in Colorado and Michigan is substantially lower and the exclusion of these individuals from earnings data is less likely to affect overall numbers. The implications of these data on commuting and migration are discussed in the conclusion in the context of barriers and challenges to cross-state data sharing.

BARRIERS AND CHALLENGES

The three projects initially supported by Data for the American Dream aim to use administrative records maintained by the state as their primary data sources. This type of work falls squarely within the allowable uses envisioned by laws governing educational data, workforce training data, and Unemployment Insurance data (with the latter being the primary source of earnings information). Adding in data from other states, however, greatly increases the complexity of the legal arrangements necessary to obtain and use administrative data. The following challenges and barriers associated with cross-state data sharing highlight this complexity:

Data sharing agreements and data governance. The primary barrier is establishing a legal relationship between the entities involved – typically state agencies – that not only fulfills obligations to the law and ensures compliance, but also sets up the ongoing governance relationships, roles, and responsibilities in ways that ensure legal, responsible, and ethical data use.

Data to support the products envisioned by Colorado, Michigan, and New Jersey could be drawn from state administrative collections on education, workforce training, and Unemployment Insurance domains. The relevant federal statutes are well known: the Family Educational Rights and Privacy Act (FERPA) and similar statutes in workforce training and UI code.¹³ This document is not intended to provide substantive legal advice, and agency staff counsel should obviously assist in drawing up the required provisions of such agreements.¹⁴

Agreements to share data across state lines must also address the governance relationships between participating entities. It is feasible to develop a data-sharing agreement that complies with relevant statutes without addressing governance questions, but the operationalization of data sharing, the ways that different stakeholders interact with one another, and the rights and responsibilities doled out and/or reserved by those agreements are generally not covered by statute and must instead be developed and agreed to by all participants.

Further, data-sharing relationships without strong governance are likely to face significant sustainability challenges, as well-developed governance regimes provide stability and continuity. A future brief will examine the issues of data governance and data-sharing agreements in greater detail.

Identity Resolution. This is another barrier to cross-state data sharing and refers to the process of determining that identities in different data systems represent the same person. It is the crucial underpinning of any successful data exchange. Presumably most of the efforts envisioned as part of Data for the American Dream will rely on matching using a combination of personally identifiable information, such as names, dates of birth, and social security numbers. Even with these data, there can be uncertainty and error in any matching process due to missing, ambiguous, or mistaken data. States engaging in cross-state data sharing must understand that there will always be some level of error involved in data matches. Samples of matched data can be examined in greater detail by all providers to get a better sense of the level of error in the matching and the potential to lead to biased results when calculating earnings associated with different credentials.



Time and effort. One of the biggest challenges to successful cross-state data sharing is the significant investment of time and effort required to establish the necessary data sharing agreements, technical infrastructure, and governance arrangements necessary to facilitate data sharing. With the ambitious goals laid out by Data for the American Dream grantees, and limited bandwidth among project teams, states may find the need to prioritize project outcomes and dedicate staff resources to the most important components of their work.

First Mover Problems. Given the extensive resources required to set up cross-state data exchanges, states that may join later stand to significantly benefit from the work done by their peers to develop necessary technical infrastructure, establish governance processes, and create data-sharing agreements. This could incentivize states to wait until exchanges are established before joining. Evidence from previous pilot projects suggests that developing a manageable cohort of states to jointly develop the processes is likely to be more successful than seeking broad immediate participation.¹⁵ In this way, the states would share the burden of start-up effort and costs.

Technology and Infrastructure. Although the technology and data infrastructure necessary to successfully match and exchange data may seem daunting to policymakers and non-technical agency staff, there are multiple potential options.¹⁶ Additionally, most states now have expertise in sharing administrative data between agencies and the technical processes are the same. Still, it is important to ensure the choices and development of a data-sharing platform meet the needs of all participants.

Data gaps in state administrative systems. Beyond missing employment data on those individuals who leave the state, Unemployment Insurance data systems have other noteworthy gaps. These systems include earnings information for all workers within the state, except for federal employees (including those serving in the military) and the self-employed. This provides broad coverage of the earnings of workers, with estimates nationally that 95 percent of workers are covered by Unemployment Insurance data systems.¹⁷ Data on federal employees had been available through the Federal Employment Data Exchange System (FEDES), but this program was suspended in 2018 and is currently being reevaluated by the U.S. Department of Labor.

While this provides broad coverage that would produce reasonable estimates of earnings, for communities with heavy federal employment or military service, or industries with significant self-employment (such as commercial fishing), this can have important implications. Additionally, increases in the so-called “gig economy” may lead to further gaps as more workers take jobs as independent contractors. Data show an increasing number of taxpayers paying the self-employment tax as new models of employment have gained traction.¹⁸

EXISTING CROSS-STATE DATA SHARING INITIATIVES AND OTHER RESOURCES

Improving data infrastructure through cross-state data sharing has taken place on many different platforms and under many different governance arrangements. This includes various bilateral efforts between neighboring states. Brief overviews of major efforts are included below:

Wage Record Interchange System (WRIS)/State Wage Interchange System (SWIS).

These two systems are facilitated by the U.S. Department of Labor primarily to address gaps in reporting on workforce programs. SWIS is the most recent version and allows participating states to query other states about earnings of program completers. Notably, the Workforce Innovation and Opportunity Act (WIOA) requires states to use data from other states to report on program performance.¹⁹ Currently, all 50 states plus Puerto Rico and the District of Columbia participate, but usage is limited to workforce programs.²⁰

Postsecondary Employment Outcomes. Through this partnership between the U.S. Census Bureau and state agencies or systems of higher education, individual education data from four states or systems have been linked to unemployment insurance records from all states and presented in aggregate.²¹ The data show flows of students between states and give a clear picture of mobility of postsecondary students. The aggregated data limits some potential uses (such as the evaluation of state or institutional programs and policies), but provides comprehensive national coverage. This effort is ongoing, with many additional states and systems scheduled to work with the U.S. Census Bureau to produce additional datasets. This work has already led to analyses of outcomes and informed policymaking.²²

Administrative Data Research Facility (ADRF). The ADRF is a secure, cloud-based platform that allows states to link administrative datasets across agencies or to other participating states. Currently, the initiative is developing a network of Midwestern states collaborating on cross-state data sharing. The ADRF is a component of a larger data infrastructure built in response to the Commission on Evidence-based Policymaking.²³

Multistate Longitudinal Data Exchange (MLDE). The MLDE is managed by the Western Interstate Commission for Higher Education and started with a four-state pilot (including Hawaii, Washington, Oregon, and Idaho) to securely share individual level education and employment data. The MLDE has recently rebuilt its infrastructure and is currently facilitating the exchange of data among a new group of states.²⁴

Other Data and Resources. Several other data resources can inform policymakers, practitioners, and researchers about the extent of cross-state mobility and its impact on data used for consumer information, policy analysis and research, and program improvement. In addition to the Postsecondary Education Outcomes resource identified above, the U.S. Census Bureau has multiple tools with publicly available data on commuting, including those used earlier in this brief to estimate state-to-state migration. These include American Community Survey commuting flow data and the Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES).²⁵ Further, the Internal Revenue Service has produced data on state-to-state migration, as well as county-to-county migration that tracks the movement of taxpayers.²⁶

CONCLUSION

With these numerous – but surmountable – barriers, and with the benefits to be gained by filling in these data gaps likely to be uneven for different states, there can be no ironclad declarations about how states should approach this topic and how much effort to commit to it. But there are several points for consideration as grantees (and others pursuing these types of efforts) work to improve opportunities for low-income, unemployed, and underemployed individuals. The conclusions below can apply equally to Data for the American Dream grantees and others working to supplement state data systems through cross-state data sharing:

- More comprehensive data on employment outcomes of program completers who find work in other states will yield better actionable information for the target audience, as well as for state agencies tasked with the responsibility of providing oversight of education and training programs.
- Agency bandwidth is not infinite. Data for the American Dream grantees have assembled impressive collaboratives of public agency staff, private entities, and others focused on improving outcomes for the population of interest. Yet, those collaboratives still have limited time and resources and there is an opportunity cost for each activity pursued. As states work to develop cross-state data sharing, staff should understand the commitment involved and work to ensure that it does not come at the expense of core activities. Still, improved data brought about through cross-state data sharing can show important differences, particularly in certain industries or programs of study and training. Previous data sharing, for example, showed statistically significant differences in salaries for completers of health-focused degrees.²⁷
- The benefits are uneven. The benefits of cross-state data sharing will vary by geography, industry, and credential type. Although labor market flow data are improving, it is not currently possible to say which credentials or jobs experience the most migration or commuting. But the data do suggest that some sectors and locations may not see much improvement from cross-state data sharing. The benefits may also accrue differently to different stakeholders. Marginally more accurate consumer information may have a limited benefit to potential students or trainees, but better outcomes information can have a substantial impact on policymaking by providing agency officials, institutional leaders, and other key decisionmakers with clear pictures of true outcomes. In the example of health programs with differential earnings, that information has significant implications for employers and institutional leaders seeking to address workforce shortages in certain areas. Demonstrating and documenting the value added by these additional data is central to sustaining cross-state data sharing initiatives.
- Progress is not rapid. The process of developing successful data-sharing relationships is likely to take significant time and may not be feasible within the initial grant period. However, the process is feasible and new data can be incorporated into systems that are initially built with a single state's data. The legal review and need to establish effective governance processes can and should be thoughtful, deliberate processes, given the sensitivity of the data involved.
- Other data resources. There are other opportunities that may be useful to improve state data resources and involve fewer obstacles and barriers (and potentially less risk). The choice of the best data pathway to follow likely depends on the specific use case.
- Key challenge remains use. In recent years, there are more and more examples of successful data partnerships that help fill gaps within a single state's data system. Doing so is not the key challenge anymore. Successfully using the data – and making an impact in the lives of low-income, unemployed, and underemployed individuals, remains the biggest hurdle for Data for the American Dream.

Endnotes

- 1 Armstrong, John, and Christina Whitfield. "The state of state postsecondary data systems: Strong foundations." State Higher Education Executive Officers Association, (2016).
- 2 Prescott, Brian T. "Beyond Borders: Understanding the Development and Mobility of Human Capital in an Age of Data-Driven Accountability. A Report on WICHE's Multistate Longitudinal Data Exchange Pilot Project." *Western Interstate Commission for Higher Education* (2014).
- 3 See U.S. Census Bureau, "Postsecondary Employment Outcomes." (2018). Retrieved from: https://lehd.ces.census.gov/data/pseo_experimental.html 24 January 2020.
- 4 Ibid.
- 5 Internal Revenue Service, "SOI Tax Stats – Migration," (2019). Retrieved from <https://www.irs.gov/statistics/soi-tax-stats-migration-data-2017-2018> 22 January 2020. The number of taxpayers counted as total number of exemptions filing income tax in home state in 2017 and a different state in 2018.
- 6 Zhang, Ting, and David W. Stevens. "Neglecting the 'L' in a Longitudinal Integrated Data System Can Be a Costly Mistake," The Jacob France Institute and the University of Baltimore, (2013).
- 7 Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. "Internal migration in the United States." *Journal of Economic Perspectives*, 25, No. 3 (2011): 173-96.
- 8 U.S. Census Bureau, "OntheMap," (n.d.) retrieved from <https://onthemap.ces.census.gov/> 23 January 2020. Author analysis of data calculated migration by earning level.
- 9 Ibid.
- 10 Author analysis of United States Census Bureau, "American Community Survey Commuting Flow Data," retrieved 20 January 2020 from <https://www.census.gov/data/tables/2015/demo/metro-micro/commuting-flows-2015.html>.
- 11 Author analysis of United States Census Bureau, "American Community Survey Commuting Flow Data," retrieved 1 June 2020 from <https://www.census.gov/data/tables/2015/demo/metro-micro/commuting-flows-2015.html>.
- 12 U.S. Census Bureau, "OntheMap."
- 13 For relevant law on exchanging wage data from Unemployment Insurance data systems please see 20 U.S.C. § 603. Additional information on regulations governing sharing of these data are summarized in joint guidance by the U.S. Department of Labor and the Department of Education, available here: https://wdr.doleta.gov/directives/attach/TEGL/TEGL_7-16.pdf.
- 14 In addition to the federal guidance mentioned in the previous footnote, further resources related to data sharing agreements covering the use of individual-level education data are available from the Privacy Technical Assistance Center (PTAC) <https://studentprivacy.ed.gov/>. An example of the federal data sharing agreement that enables the State Wage Interchange System (SWIS) is available here: https://www.doleta.gov/performance/pdf/SWIS_Agreement_6-20-19_Accessible_PDF_Fillable_Form.pdf
- 15 Prescott, 2014.
- 16 See for example the Administrative Data Research Facility (ADRF) <https://coleridgeinitiative.org/> and the Multistate Longitudinal Data Exchange <https://www.wiche.edu/longitudinalDataExchange>.
- 17 U.S. Bureau of Labor Statistics, "Quarterly Census of Employment and Wages," (2018). Retrieved from <https://www.bls.gov/cew/overview.htm> 20 January 2020.
- 18 Duca, John V. "Inflation and the Gig Economy: Have the Rise of Online Retailing and Self-Employment Disrupted the Phillips Curve?" Federal Reserve Bank of Dallas, Working Paper 1814, (2018).
- 19 United States Department of Labor, "State Wage Interchange System." Retrieved 1 June 2020 from <https://www.dol.gov/agencies/eta/performance/swis>.
- 20 Ibid.
- 21 United States Census Bureau. "Post-Secondary Employment Outcomes." Retrieved 1 June 2020 from https://lehd.ces.census.gov/data/pseo_experimental.html.
- 22 See for example: Huie, Stephanie Bond and David R. Troutman. "A Roadmap to Better Data: Developing a Census Bureau Partnership to Measure National Postsecondary Earnings Outcomes." Washington, DC: Institute for Higher Education Policy, 2019. Retrieved 1 June 2020 from http://www.ihep.org/sites/default/files/uploads/docs/pubs/ihep_ut_census_v3_0.pdf
- 23 Coleridge Initiative. "ADRF". Retrieved 1 June 2020 from <https://coleridgeinitiative.org/adrfl>.
- 24 Western Interstate Commission for Higher Education (WICHE). "Multistate Longitudinal Data Exchange Overview." Retrieved 1 June 2020 from <https://www.wiche.edu/longitudinalDataExchange>.
- 25 Interested readers seeking details on the differences between these two tools should consult Graham, Matthew R., Mark J. Kutzbach, and Brian McKenzie, "Design Comparison of LODES and ACS Commuting Data." Washington, DC: United States Census Bureau, 2014. Retrieved 1 June 2020 from <https://pdfs.semanticscholar.org/77e2/a4c8fae3c1affa0aed646aab693a26fb2f9c.pdf>
- 26 Internal Revenue Service. "SOI Tax Stats." Retrieved 1 June 2020 from <https://www.irs.gov/statistics/soi-tax-stats-migration-data>.
- 27 Prescott, Brian. "Beyond Borders: Understanding the Development and Mobility of Human Capital in an Age of Data-Driven Accountability." Boulder, CO: Western Interstate Commission for Higher Education, 2014. Retrieved 1 June 2020 from <https://www.wiche.edu/content/beyond-borders-understanding-development-and-mobility-human-capital-age-data-driven>.

D4
AD

WWW.D4AD.COM

