

Fostering State-to-State Data Exchanges

BY BRIAN T. PRESCOTT

NATIONAL CENTER FOR HIGHER EDUCATION MANAGEMENT SYSTEMS

PATRICK LANE

WESTERN INTERSTATE COMMISSION FOR HIGHER EDUCATION

MAY 2016

Brian T. Prescott, Ph.D., is associate vice president of the National Center for Higher Education Management Systems. Previously, he led the development of the Western Interstate Commission for Higher Education's Multistate Longitudinal Data Exchange from its inception until he left WICHE in November 2015. **Patrick Lane**, Ph.D., is project manager for WICHE's Multistate Longitudinal Data Exchange.

This paper is part of the larger series *Envisioning the National Postsecondary Data Infrastructure in the 21st Century*. In August 2015, the Institute for Higher Education Policy (IHEP) first convened a working group of national postsecondary data experts to discuss ways to move forward a set of emerging options for improving the quality of the data infrastructure in order to inform state and federal policy conversations. The resulting paper series presents targeted recommendations, with explicit attention to related technical, resource, and policy considerations. This paper is based on research funded in part by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the author(s) and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation or the Institute for Higher Education Policy.

Executive Summary

Gaps in information available from the nation's current data infrastructure about postsecondary education outcomes, including employment, have left policymakers inadequately informed about institutional performance and the effectiveness of the policies they put in place. These gaps also leave prospective students and their families in the dark about how best to make their educational investments pay off as they make life-altering decisions. Moreover, the gaps complicate the use of evidence to inform improvements in policies and practices that impact student success.

One potential solution that addresses many of these limitations is to forge linkages between existing state data systems in order to meet the differing education and employment data needs of diverse and varied stakeholders, including students and their families, institutions of higher education, workforce training providers, state policymakers, and the federal government. Such an exchange can be set up to be safe and secure and maintain privacy, using a limited set of data elements to unlock a huge range of potential analytical opportunities. Once in place, it could help states hold their postsecondary institutions (and workforce training providers) accountable; evaluate and improve policies and programs; provide more complete information to prospective students about their likelihood of succeeding in postsecondary education or training, and securing a paying job; and inform statewide and institutional strategic planning efforts.

Drawing from the experience the Western Interstate Commission for Higher Education (WICHE) has acquired in leading the Multistate Longitudinal Data Exchange, this paper describes how such a system might work, how it complements a more contemporary and effective data infrastructure for the nation, and what challenges must be addressed and surmounted for the vision of a sustainable data exchange among states to be realized. Lessons from the MLDE to date show that a cross-state data exchange is possible, that it provides additional information useful to the states that participate, and that, by leveraging federal and state investments already made, it can operate at a manageable ongoing cost. These elements provide the foundation on which to build a state data exchange, and it would be built with states' data needs foremost in mind.

But the experience to date shows that research using state data systems has led to new information about the impact of state policies, such as financial aid and remediation, and those insights have also featured prominently in national policy debates. A state-based resource that better captures

student outcomes—by making it possible for states to flexibly share individual-level data—can substantially build on that growing body of research while improving results. Even more, a state data exchange can provide states with the evidence base they need to operate institutions more efficiently and effectively, hold them accountable, address equity goals, improve policies and programs, and adopt and monitor strategic plans.

No matter the intended use for state and federal policymakers or institutional leaders, however, the most significant challenge faced by a state-to-state data exchange is incomplete geographic and institutional coverage arising out of the exchange's reliance on voluntary state participation. This lack of complete coverage constrains its utility for some purposes, such as holding private nonprofit and for-profit institutions accountable, particularly for federal financial aid programs. Getting to the point where all states and all students are accounted for will not happen overnight. Even well before it achieves that aim, a state data exchange will need to establish a sustainable state-led governance structure and business model that ensure both adequate voice and representation from all participating sectors (e.g., K–12 education, public and private postsecondary education, and labor market information providers), as well as adequate ongoing funding support.

Fortunately, the project has made substantial headway in addressing technical issues related to data sharing. It has done this in large part by adopting a federated data model that ensures participating states have maximum flexibility to obtain the data they need while incorporating robust protections for data security and individual privacy. As the project evolves, it will continue to adjust the configuration of the procedures to match records across states and systems. It is critical that the final approach is transparent to participating states. Once operational, states will have to “plug in,” but data pathways and data definitions are being standardized to help with those efforts.

A second important challenge to address as a state data exchange matures is the capacity within state agencies to focus on cross-state data and analyze data as they come in. States vary considerably in how much time and energy they have available to commit to longitudinal data systems' development, management, and use. (John Armstrong and Katie Zaback's paper, *Assessing and Improving State Postsecondary Data Systems*, provides an overview of state postsecond-

ary data systems.) A fully realized state data exchange will be able to provide data products and tools not just to help participating states, but also to enlist other analysts and researchers in ensuring effective use. WICHE's MLDE is expected to resume exchanging data within the coming year. After a suitable period for participating states to analyze the MLDE's products, they will start to wrestle with how and under what conditions to permit access to analyst contractors.

WICHE's MLDE project is currently in a second phase that will conclude by summer 2018 with expansion of the exchange to serve at least 10 participating states. By then, its version of a state data exchange is expected to be operational and both self-governed and self-sustaining by the participating states. Interest among states across the country has been high, however, and future prospects for the project improve as more states agree to join and share the costs of ongoing operations. Based on cost estimates that include relatively firm future expenses for software licensing and data storage needs, plus reasonable estimates for expenses related to governance, the MLDE projects to run \$500,000–\$600,000 annually. Because the costs of handling and storing data are so low, additional states will add only marginally to those estimates.

As at least a partial solution to bringing the nation's postsecondary data infrastructure in line with current and future needs, a state data exchange comes with intriguing political dimensions. Standing in its way is the need to attract individual states to the exchange by convincing them of its value and assuring them that its architecture and governance components will protect sensitive data and individual privacy. Meeting this challenge is not an insignificant task. However, with growing demands for flexible access to data for use in post-collegiate and employment outcomes along with support from state-based data architects, it has been manageable so far (albeit with less success among K–12 education agencies) in the MLDE. By contrast, a state data exchange can make significant headway without running afoul of two powerful interest groups that can be counted on to resist a federal solution: those who would claim federal overreach if the federal government issued a mandate that individual-level data on all students be collected by the federal government and those in the private-school lobby who have previously been resolutely opposed to a federal unit-record system. A state-based voluntary system may be an appealing alternative.

To advance the concept and value of a state data exchange, this paper includes several recommendations for federal and state policymakers and agency leaders. Those applicable recommendations for federal policymakers and officials include the following:

1. **Elevate the priority the federal government places on cross-state data partnerships.** This can be accomplished two ways. The first is to require cross-state data to be an essential part of any future competitions for the Statewide Longitudinal Data System and Workforce Data Quality Initiative programs. The second way to elevate cross-state data partnerships is for the federal government simply to mandate the use of a cross-state data exchange as part of the Higher Education Act or some other legislation. Such a step would instantly address the weakness of incomplete coverage of a voluntary data exchange, but it might be counterproductive if such a mandate is perceived to be an instance of the federal government overstepping its authority.
2. **Fund analytical work that demonstrates how mobile our society has become and the importance of accounting for such mobility in examining educational outcomes.** As the effort to examine employment outcomes for recent college graduates builds momentum, the federal government should fund more work that demonstrates how mobile our society has become and how that mobility must be accounted for in drawing conclusions about the value of various kinds of postsecondary credentials. Doing so with data linking education and employment records longitudinally would help flesh out how individuals with different academic and vocational backgrounds are finding success, or not, over a longer period than the current focus on short-term earnings permits and with a richer evidence base than snapshots taken at specified intervals provide.
3. **Make it possible for states to access and use information housed in federal data systems in order to more comprehensively and accurately analyze employment outcomes and evaluate educational policies and practices.** The federal government can work with states to provide access to the postcollegiate employment experiences they cannot access on their own, namely self-employment and federal and military employment, so that states have all the information they need to manage their institutions and provide useful information to consumers.
4. **Require submission of additional data elements to unemployment insurance wage records.** Expanding the minimum requirement of what must be reported in state's UI system by the addition of three data elements would greatly improve the information state data systems are able to provide to policymakers and the general public. Those elements are occupation, hours worked, and location of workplace.

Recommendations for state policymakers or agency leaders include the following:

- 1. Demand information that captures the outcomes of all students by joining a state data exchange like the MLDE and by using existing state data sharing resources.** State policymakers can take care to not be fully satisfied with incomplete information about postcollegiate outcomes that does not account for out-of-state mobility for employment or further schooling. One way they can do so is by insisting that their state be engaged in a state data exchange project like WICHE's MLDE.
- 2. Ensure that state data systems include elements needed to produce and use disaggregated results to improve equity in education and employment outcomes by program and for specific population groups.** State policymakers should note that wide disparities in institutional mission and program mix mean that aggregate data on postcollegiate outcomes at the institutional level may be misleading while also obscuring important differences in the experience of different groups of students. These conditions highlight the need for disaggregated data to ensure that they and the institutions themselves have the tools to make policy and program improvements.
- 3. Require additional data elements be submitted as part of employers' UI data submissions.** As demonstrated by states like Nebraska and Louisiana, which are asking employers to voluntarily submit new data, states do not have to wait for the federal government to add elements like occupation, hours worked, and workplace location to their data systems.
- 4. Ensure that data elements necessary to connect education and employment records are available for use.** State policymakers must recognize that they cannot obtain analyses of linked education and employment data without using Social Security Numbers. If they want information about the payoff to education, particularly for students who do not go to college, they can resist putting unnecessary restrictions on the collection of SSNs without jeopardizing data security and privacy, or facilitating linkages with data sources, such as those of state departments of motor vehicles, where SSNs are already collected.
- 5. Allow institutions to use employment data from state UI wage records at the unit level.** To foster the use of employment data for policy and program improvement, states can permit institutions to use unit-record employment data from the UI wage record files in a de-identified format.

Fostering State-to-State Data Exchanges

Introduction

Few topics in postsecondary policy have received greater attention in recent years by policymakers, funders, and research and advocacy organizations than the creation and use of performance measures for student success in and after college and the development of data systems to support them. This interest arises from a widely shared awareness that the nation's existing data infrastructure does not adequately meet the needs of public policymakers, students and their families, institutional leaders, or the general public.¹ In particular, a lack of individual-level data that capture the success (or lack thereof) of students who do not attend full time or who move between postsecondary institutions, as well as information about employment outcomes for all students, leaves policymakers inadequately informed about institutional performance and the effectiveness of the policies they put in place. At the same time, prospective students and their families are in the dark about how best to make their educational investments pay off as they make life-altering decisions. For their part, institutions have a huge supply of data about the students they serve, but they too face limitations in fully deploying those data for student success initiatives because they are unable to capture complete information on their students' outcomes.

One potential solution to address many of the shortcomings of the existing data infrastructure is to forge linkages between existing state data systems. This paper will discuss the potential benefits of such a system and address challenges to its implementation, drawing heavily from the experience the Western Interstate Commission for Higher Education (WICHE) has acquired in leading its Multistate Longitudinal Data Exchange (MLDE) project. (An overview of that project is available in **Sidebar 1**.) It will argue that the federal government's data needs are not synonymous with what states need or could benefit from, and that neither another federal data collection nor a data system cobbled together by states with their needs in mind is likely to fully address all the specific requirements of the other. One of the most significant ways in which these needs differ stems from states' constitutional responsibility for overseeing—and sometimes operating—educational institutions, a

responsibility that falls outside the purview of the federal government. Combined with highly variable approaches states have taken in the provision of education to their residents, states' data needs go beyond what is necessary to reliably calculate various performance metrics to enable them to appropriately gather individual-level data for use in operational decision-making (e.g., program approval); analyses of institutional or program performance (e.g., evaluating a statewide program); and statewide strategic planning (e.g., tuition pricing, designing economic development initiatives). The paper will also discuss how state-to-state data linkages can be set up to be safe and secure while respecting privacy, and outline ways that the exchange of a limited set of data elements can unlock a huge range of potential analytical opportunities and applications or uses. Along the way, the paper will acknowledge the challenges being addressed in cross-state data sharing, and describe the MLDE approach as well as cover alternative solutions.

Background and Context

Even though the U.S. Department of Education (ED) is limited by the ban on a federal unit-record data system in the Higher Education Opportunity Act of 2008, it has been actively pursuing ways it can address the demand for data to better capture student success in achieving a credential and a well-paying job. With the recent release of the College Scorecard, ED took what steps it could to release information on student outcomes, including earnings, drawn from administrative data at the individual level, although the results were for financial aid recipients only. Another way ED has tried to meet the demand for better data is by making heavy investments in state efforts to advance the use of data through the Statewide Longitudinal Data System (SLDS) grant program. Together with the Workforce Data Quality Initiative (WDQI) run by the U.S. Department of Labor, federal spending has exceeded \$750 million to 48 states since 2005. When combined with what the states themselves have contributed, these investments provide the foundation for the state data exchange—a utility or shared resource through which states share data with one another about the students their educational institutions have served.

SIDEBAR 1: WICHE'S MULTISTATE LONGITUDINAL DATA EXCHANGE

Beginning in 2010, WICHE convened representatives from four states—Hawaii, Idaho, Oregon, and Washington—to test whether it was possible to exchange individual-level data among them and, if so, how much new information such an exchange would reveal.² By mid-2014, the four states had successfully exchanged individual-level student data on more than 190,000 students, with each state receiving enhanced data in identified form for the individuals it contributed to the MLDE in the first place. This initial exchange showed that participating states were able to fill between 9 and 22 percent of employment data missing from their own systems.³ It was also possible to detect differences in the wages earned by individuals who stayed in the state following graduation and those who moved away, a finding that calls into question the validity of assumptions that are necessary when relying solely on one state's data.⁴ States also realized how data combinations could improve analytical approaches. For instance, summing wages earned during the same quarter over multiple states could capture a truer measure of earnings for workers who switched jobs during the quarter or who worked two jobs concurrently on either side of a border. Also, by combining employment data with data on subsequent enrollment supplied by both participating states and the National Student Clearinghouse, the analysis of earnings could directly account for former students who were both employed and enrolled, as well as crediting students who elected to pursue further education but for whom no employment records could be found.

The MLDE pilot also brought forward two unsurprising but nevertheless unanticipated lessons about the importance of mobility in understanding how states could be keeping track of the payoff of their educational investments, as well as how entangled with neighboring states those investments inevitably become. The initial insight came from the realization that the MLDE dataset provided WICHE and participating states with a unique view of how former students move as they exit postsecondary education. Such mobility—from high school to college and after leaving college (with or without a credential)—varies by state and academic program. The second important realization was that the states could obtain information about the extent to which they were able to attract graduates from nearby states to their labor markets.⁵ Heretofore, state policymakers have had no tools to evaluate the extent to which the mobility of individuals as they leave college affects their labor market.⁶ In other words, the MLDE can uniquely supply insights about the “balance of trade” in human capital among states—how information about the flow of recently educated individuals into labor markets empowers policymakers and institutional leaders to better understand their alignment to the specific labor market needs of their states. Moreover, since neither macroeconomic conditions nor industries nor firms stay neatly within the confines of state borders, states using the MLDE will have a more realistic picture of how labor markets shape up in regions. This can lead to a more productive and coordinated multistate strategy for ensuring that the combined educational investments of states add up to the labor market need. This is an arguably more efficient and productive path than for individual states to be going it alone.

States have adopted different priorities and strategies for how they are developing and using the data systems they and the federal government have funded, but the progress they have made is paying dividends in state policy decision-making.⁷ Perhaps the most direct evidence of this can be seen in sophisticated, data-driven outcomes-based funding policies that are tying institutional funding levels to student success metrics. Two states, Florida and Texas, have even gone so far as to implement measures of employment outcomes as part of such policies.⁸ More commonly, states are also beginning to publicly release information on student outcomes to help prospective students make informed choices—often as required by state legislation.⁹

The demand for consumer information about employment and wage outcomes is so powerful that nongovernment entities are getting into the game. LinkedIn, PayScale, and Burning Glass are perhaps the three most obvious examples of corporations using their own proprietary data to weigh in with estimates on how much graduates of different programs at different institutions are able to earn.

Yet accountability and consumer information are not the only worthwhile uses of a better postsecondary data infrastructure that captures educational and employment outcomes. Less apparent but no less important is that institutions also suffer from a lack of information they can use to compare their performance with that of other institutions. Also, state policymakers are limited in their own ability to judge the best formulation of policies to align investments and incentives with statewide goals for educational attainment, affordability, or productivity. Better information is needed to support policy and practice improvements at the federal, state, and institutional levels, and to plan strategically for an uncertain future.

There is no question that the federal government needs information that allows it to effectively steward taxpayer funds that make their way to postsecondary institutions through financial aid programs and to provide consumer information. These have been expressed as the primary purposes for federal data collection by the U.S. Senate's Committee on Health, Education, Labor, and Pensions in its summer 2015 position paper on the postsecondary data systems: “During the upcoming reauthorization of the Higher Education Act, policymakers have the opportunity to refocus the scope of data and disclosure policy... [to] promote purposeful and accurate data for evaluating the efficacy of federal student aid programs and providing transparency to students and families on postsecondary options.”¹⁰ Through the Integrated Postsecondary Education Data System (IPEDS) and national sample surveys, the federal government also has assumed a lion's share of the responsibility

ity for ensuring that data are publicly available to stakeholders, whether that comes in the form of consumer information or as datasets to be used for the production of accountability metrics or for research. Furthermore, the federal government has access to other data sources only now beginning to be tapped for analyzing the effectiveness of postsecondary education, like taxpayer records linked to federal aid recipients in the College Scorecard to produce aggregate statistics. These capacities are crucial for administering federal financial aid programs and for holding institutions accountable for the dollars they receive from those programs, and for providing the empirical foundation for what we know about how higher education is performing.

However, as the entities that own and operate public institutions, which collectively enroll about three-quarters of all students nationally, states require a data infrastructure geared toward their needs as well. Yet their data needs and those of the federal government are not identical, although each stands to benefit from progress in data system development and usage by the other. State policies are more directly influential over institutional decision-making in the public sector. This is because states typically oversee everything from the operation of the institutions themselves to the academic programs institutions are approved to offer, to admissions policies, to developmental education, to how well and through what means institutions are funded to meet their respective missions. States' needs also extend to the operational oversight of institutions. The data needed to properly evaluate the impact of decisions along all these dimensions are different in scope and scale than what are needed to satisfy the federal government's chief aims for postsecondary data. Moreover, states need specific information about how success in college may be related to prior preparation in high school and to the future workforce and economic development needs of specific regional and statewide economies. Accordingly, states' needs for data linkages at the individual level span the intersections between K-12 education, workforce development programs, employment, and even other human services programs.

Still, in designing and implementing a data system to capture all relevant information on the individuals they have served educationally, states are at a big disadvantage: Their borders tend to be a significant hurdle to the movement of individual-level data even as they pose no real barrier to individuals—or, for that matter, to industries and firms that might be seeking to hire or to macroeconomic conditions that will influence the opportunities that are available to individuals and firms across a multistate region. Research clearly illustrates how essential a feature of American life mobility is, as well as how connected it is to educational attainment.¹¹ The net result is that, over time, state data systems will capture an unpredictable, constantly shifting, and gradually diminishing group of individuals.¹² These changes have an unknown

impact on the results of any analysis of educational and employment outcomes. However, it is quite plausible that Americans' tendency to be mobile will skew any findings if left unaccounted for, especially in regions of the country that feature many small states, or for certain academic programs that are located at a distance from where related jobs tend to be clustered.

Forging linkages between state data systems can begin to fill these gaps. A natural complement to what the states are already developing, these linkages—essentially a data utility that allows for sharing cross-state and cross-sector data—would also provide a means to unlock the full potential of unit-level data to inform policy and practice, not just to support accountability and consumer information. Such are the challenges WICHE's MLDE is being developed to address. Now in its second phase, the MLDE has already demonstrated that states can and will link records at the individual level, that participating states can obtain information on subsequent education and employment outcomes of the students they serve who cross state lines, and that such a resource is valuable and possibly necessary to evaluate and improve policies that are instrumental in how human capital is developed and deployed. Among the most valuable lessons learned thus far is how states can link their existing data together to help better address the demands for accountability and consumer information, without allowing those needs to overshadow the critical role states and institutions can play by using their data for continual improvement and planning. The value of a state data exchange grows as more states become involved, so it figures to be most effective if all 50 states participate. But even short of that goal—after all, getting to universal coverage will take some time—the additional information states can obtain by collaborating is substantial.

WICHE's experience suggests that an optimal postsecondary data infrastructure must go beyond the need for better metrics more accurately calculated for more students. Those metrics are of tremendous value for assessing progress toward state or national goals and for institutions to use as benchmarks. But to account for all the variation in state policies and programs that affect educational success—and economic development more broadly—states need flexible access to linked individual-level data that cross states and sectors. A stand-alone data collection that, once assembled, is thereafter limited to the data elements, covered entities, and time boundaries contained within would not provide for the customization that individual states could use to evaluate and improve their own unique policies, programs, and conditions. As a result, states are unlikely to respond to the presence of an improved data system housed by the federal government by shuttering their own data systems. Nor is the demand among states for identifiable cross-state data likely to dissipate.

How Would a Multistate Longitudinal Data Exchange Function?

Broadly speaking, there are three important components to consider in stitching states' data systems together: the architecture by which the data exchange is organized, the governance under which it operates, and a business model that ensures it has the financial means to sustain itself. This section focuses on the architecture, to provide an overview of how a data exchange might work. The architecture addresses the overall design that allows data to pass among participating entities and be stored, how the rules established by governance are adhered to, and the detailed technological solutions needed to implement the data exchange's record-matching and data transmission and storage processes. Issues of governance and sustainability are vital and will be addressed later in this paper. However, it is worth noting that governance is interwoven into the design and throughout the data sharing agreements that are at the heart of the exchange's architecture.

Mirroring the different approaches states have taken in constructing their own SLDSs, there are two approaches to linking states' data systems together. (See **Sidebar 2** for a discussion about the role of unemployment insurance [UI] wage data in a state data exchange.) In the first approach, known

as a warehouse model, states would submit files of individual records containing a set of data elements they have collectively agreed on to a warehouse in which the records would be matched. Next, the warehouse operator would request participating states to submit additional data for any students whose records were initially provided by any of the other participating states, which it would then append to individual records in the warehouse. The result would be one dataset (or a relational database) containing all the information available across multiple states for the cohort(s) of students selected for inclusion by the participating states at the outset. The warehouse operator would prepare datasets for each participating state that included only those students the state was permitted (legally and according to the exchange's data sharing agreements) to receive information about—generally those students the state submitted to the data exchange in the first place. A de-identified dataset would also be created for regional or national analyses, with access to that data restricted by the exchange's data sharing agreements. However, at minimum, an appropriate third-party—such as the entity that negotiated the data sharing agreements—would be given responsibility for analyzing those data. The warehouse model is the approach WICHE took in its initial data exchange effort between 2010 and 2014.

SIDEBAR 2. WHY RELY ON UNEMPLOYMENT INSURANCE WAGE RECORDS?

A number of options exist for capturing employment and wages and using those data in indicators designed to measure the return on educational investments, the likelihood of finding a job, and so on.¹⁶ Of these, state UI wage files, while extremely useful, are often characterized as the most limited, with the main weaknesses being that they cannot provide any information about self-employment or employment in the federal government or the military. They also apply to just one state at a time, except for aggregate figures from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program and efforts to link UI wage records across states. So why not use a more comprehensive source for a state data exchange?

The answer is twofold. First and most importantly, UI wage records are owned by the states, which means the states can set the rules under which they are willing to exchange those records with one another and how the records are subsequently used. In particular, under very carefully shaped agreements, states have shown a willingness to share individual-level UI wage records with one another, like through the Wage Record Interchange System (WRIS) and the pilot phase of the MLDE. The ability for states to gather individual-level wage data is a major advantage over alternative data sources run by the federal government that generally allow data to be released only in aggregate. A state data exchange model like the MLDE needs individual-level data for several enhanced uses, including the following:

- ▶ Examining the extent to which students are retained as employees in the states where they were educated

- ▶ Obtaining a better measure of total wages by summing wages recorded for the same individual in the same or in adjacent quarters in multiple states
- ▶ Disaggregating employment outcomes for academic and vocational programs and for special populations (like underrepresented minorities, state financial aid recipients, students with experience in remedial education, noncompleters, etc.)
- ▶ Reporting on variation in employment outcomes like wages among graduates, both to represent the distribution of wages and to examine differences based on where former students find work
- ▶ Evaluating policies for which having identified employment records would be useful in linking to other data elements, such as how well students with debt are able to afford their repayment obligations, or how likely state grant recipients are to stay in state following degree completion
- ▶ Making possible the construction of a longitudinal dataset that is not simply a series of snapshots of a state's labor force participants without accounting for swirling patterns of enrollment and employment that include mobility across state lines (which David Stevens and Ting Zhang demonstrate can be substantial¹⁷)

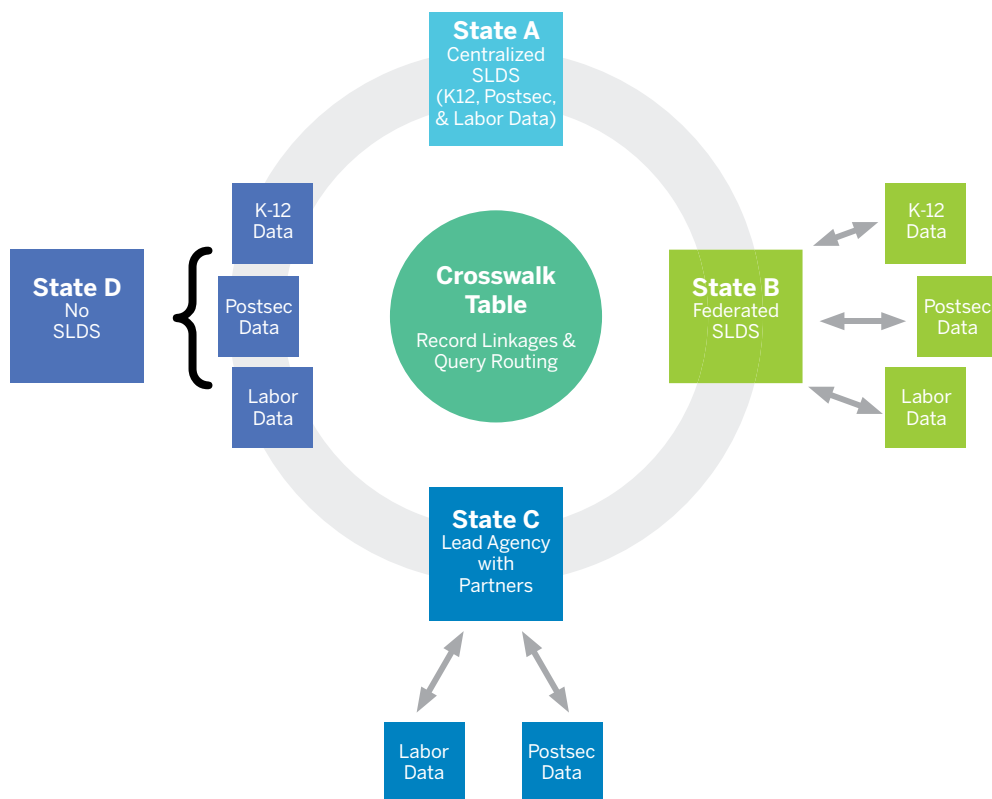
Finally, relying on UI wage records does not preclude states individually, or collectively through a state data exchange like the MLDE, from supplementing them by gathering aggregate data from other sources that cover employment not captured by UI wage records.

The alternative approach is to federate the data, which is the architectural design adopted by the MLDE project going forward. In this approach, only the most parsimonious amount of data is to be held centrally—namely those elements necessary to adequately resolve identities and link records across states and data systems. Participating states then exchange the bulk of the data directly with one another only upon receipt of a valid request. **Figure 1** illustrates the concept. At the center are the key functions needed to make the linkages between records shared by participating states and routing queries among them. With respect to the latter, and as further described below, states would share most information directly with one another—not with the central hub, which would simply process the incoming query by passing it to the remaining states along with information about what data are being requested on which students and where to send the resulting matches. Under this architectural design, each state might be viewed as a “node” in a network of states linked together. Additional states that agree to participate would simply come aboard as another node in that network. As depicted in the figure, participating states organized under a diverse range of data governance models can be accommodated. This allows the exchange to include states that operate a tightly coordinated central SLDS as well as states that have no shared data system at all, and everything in between. There is no technical limit to the number of nodes the architecture can accom-

modate, so all willing states could participate. In fact, other, nonstate data sources could also be considered nodes on this network, albeit with some adjustments to how the data are exchanged and at what level of granularity. In particular, nonstate sources would simply be suppliers of data, unable to access the core record-matching or query processes. Specific procedures for accessing and limits on using data from nonstate sources could be set up in accordance with those sources’ requirements. For example, the exchange might be an efficient way to submit queries on behalf of multiple states to a data source like the Federal Employment Data Exchange System (FEDES) that has strict rules about how and to whom it can supply responses. Or states could collectively request aggregations of data about their former students’ employment in non-participating states from the Wage Record Interchange System (WRIS and WRIS2), the U.S. Census’s Longitudinal Employer-Household Dynamics program (LEHD), or the Social Security Administration (SSA).

Figure 2 details how this model’s central functions will work. First, states provide a limited set of data elements sufficient to match records once or twice annually.¹³ The records to be included would cover all students in postsecondary education in whatever years states are collectively willing to provide, not limited to cohort definitions (e.g., beginning postsecondary students to be followed in subsequent years) as with the warehouse model. The matched records would be assembled into a “crosswalk table” consisting of participating states’ purpose-built identifiers linked where matches exist. Participating states would then query the crosswalk table using their identifier, which would generate a request for other states to send corresponding enrollment, awards, and employment information directly to the querying state. In this model, no data leave the state apart from what are necessary to make matches, until the states receive a properly vetted request from the exchange’s automated system. Further securing the data and protecting privacy is a provision that the personally identifiable data used to make the record linkages and build the crosswalk table will be held offline, inaccessible without physical access to the servers on which those

FIGURE 1. FEDERATED STATE DATA EXCHANGE MODEL



data rest.¹⁴ As with their own unit-record data, participating states would expect one another to adhere to a robust set of security standards. This approach means that no central resource exists where all data are present together in personally identifiable form. But it permits states to construct queries that fit their specific policy questions.

Using this model, individual states would be able to calculate metrics related to performance, efficiency, and equity by supplementing the data elements they keep in their data systems with comparable data elements from other states. Such calculations would be more accurate than the states could make on their own because the addition of other states' data would make the data they have at hand more comprehensive. Additionally, states would be able to use these metrics along with the flexible querying capacity provided by the data exchange to understand and improve their varied policies and programs. And like the warehouse model, the federated approach allows for the construction of de-identified datasets for the calculation of such metrics across multiple states. To do so, states would submit queries on a commonly defined cohort of students through the exchange. They would then de-identify the results before sending those files along to a third-party entity working under contract to produce an approved analysis. That entity would build the de-identified data file using the one-time-use, query-specific identifier generated by the exchange. For most metrics, the de-identified dataset would be sufficiently comprehensive for answering the most common questions about student success in postsecondary education and the workforce, such as what is an institution's graduation rate, how many of its graduates were able to find employment, and with what level of earnings?¹⁵

A Multistate Data Exchange's Role in a Contemporary Postsecondary Data Infrastructure

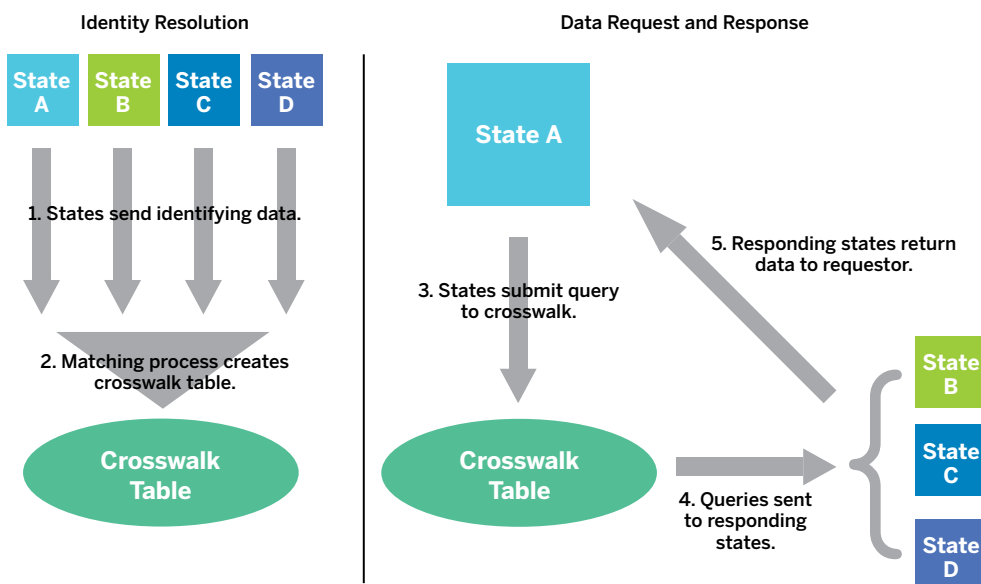
Among the options being put forward to improve the nation's postsecondary data, a system that links state data systems together occupies something of a unique position. A first distinction is that a state-based data exchange would be intentionally aimed at state policymakers and institutional leaders far more than at federal policymakers. Since states both operate public institutions and set the regulatory and policy context for private institutions within their borders¹⁸—not to mention house the varied economic development conditions students will encounter upon leaving their studies (whether they graduated or not)—they need data that cover a wide and changing array of policy topics that are instrumental in determining educational outcomes and affordability. States can benefit from having ready access to the kind of broad performance metrics at the heart of the push for better national postsecondary data. However, it would be helpful to have information about whether former students are employed or continuing their studies in the same state or elsewhere, as well as estimations of how much burden students with debt are taking on in repayment (by, for instance, putting their debt obligations over their earnings).¹⁹

But there are alternatives to the effort that creating and maintaining data linkages spanning multiple states requires, which can help provide much of the data needed for such metrics. The most obvious existing option for linking to education records is the National Student Clearinghouse (NSC), which already offers near-universal coverage of student enrollment and awards in postsecondary education and with which many states already gather data.²⁰ Indeed, the NSC was a key partner on the MLDE's first phase,

where its data complemented what the pilot states exchanged in order to capture students' enrollments in private institutions as well as public institutions located outside the four states. A state data exchange solution would almost certainly seek to continue that complementary partnership.²¹

Regarding wage data, a number of options are available, as outlined by Rachel Zinn in *Classroom to Career: Leveraging Employment Data to Measure Labor Market Outcomes*.²² Applied to the consideration of a state data exchange, her chart is missing only one key attribute related to whether state educa-

FIGURE 2. RECORD LINKAGE AND QUERY PROCESS



tional agencies can obtain unit-level data, or just aggregated results. Of the data sources Zinn lays out, leveraging data on employment and wages hosted by the federal government (such as the LEHD program or the SSA) or tapping into the WRIS are the two most obvious ones. While each of these sources offers states the capability to link to employment outcomes data in all states, they will provide only aggregate results. These broad measures would make it possible to address some fairly straightforward questions related to accountability and consumer information, such as calculating a median wage for a graduating class. It would also be possible to provide some basic information relevant for policy and program improvement via one or another of these alternatives, at least insofar as aggregate information might be helpful for that purpose.²³ States should take advantage of such resources where possible and when they can supply data sufficient to answer the analytical questions at hand.

Nevertheless, a state data exchange can help provide otherwise unavailable information for use in accountability systems and in providing reliable information for consumers. Take, for example, the outcomes-based performance funding policy used by Florida's public four-year institutions.²⁴ The first two performance measures reward institutions for how well their graduates are able to find employment and what their median wages are. For the first measure, Florida uses WRIS2 to supplement what it knows about how many of its graduates found jobs in Florida with information about job placement in other states. But for the second measure, WRIS2 is unhelpful since the aggregate data it provides cannot be combined effectively with the unit-level data in Florida to produce medians.²⁵ Additionally, both measures are produced for the institution, while real accountability related to employment outcomes may more properly be tied to academic or vocational programs.²⁶ Gathering that data through WRIS2 would likely be either impossible or at least extraordinarily burdensome to the state agencies expected to compile those data in aggregate form, although loosening that restriction to allow education agencies to receive unit-level data from WRIS2 would improve the chances that it would be done by letting the agencies that need that information take on the task of producing it. WRIS2 covers only the two most recent years, which also limits its usefulness by preventing longitudinal analysis over a longer time horizon. For instance, except for very short-term programs, it is not possible with WRIS2 to estimate the value added to individuals' median wages based on the educational program they undertook: WRIS2 would permit analysis of wages before enrollment or after completion, but not both.

These limitations also pertain to the quality of consumer information that can be produced, though perhaps less acutely given the lower stakes associated with that use relative to accountability. Still, if an implicit goal for producing

consumer information is to equip prospective students with information that is as accurate as possible, median wage figures should include data on the wages of former students who find work in another state where they generally earn more (or less)—as WICHE found in its MLDE pilot phase—or as one might hypothesize for certain majors for which the target job market tends to be clustered geographically, like a business program that sends a large number of graduates to Wall Street. Consumer information might also be enhanced by the ability to conduct truly longitudinal analyses of wage and employment outcomes, where the same students are tracked over time (not simply snapshots of what subset of a cohort happens to appear in wage records at the time the metric is calculated), or the ability to provide some sense of the wage variation a student might expect, to supplement the published median wage. A state data exchange could provide these more comprehensive views of outcomes that would be difficult or impossible to produce with alternatives that do not provide unit-level data back to states.

The capability to provide this additional information boils down to the flexibility and utility that come with forging linkages between states' data systems and to improvements in data use and potentially quality. First, a state data exchange (operating under a federated model) is not so much a data collection (like, for example, IPEDS, national sample surveys, a hypothetical federal unit-record system, or even the warehouse data exchange model created during the initial pilot phase for MLDE) than a shared resource that maps the relationships between existing data residing in multiple states and provides mechanisms to exchange a limited set of relatively well-understood data elements. With these elements providing a map of the linkages between data, states are able to use the MLDE to gather key pieces of information they would otherwise lack, and to combine the exchanged data in meaningful ways with other data they have on hand about students' educational experiences, to use in a wide array of analytical applications.

Therefore, in a state data exchange, participating states' data are not sidelined by prematurely forcing the exchanged records to be de-identified. There is no question that de-identified data are adequate to address many policy- and practice-relevant questions, while not all analysts need access to identifiable unit-record data and some have no need for unit-record level data at all.²⁷ But the timing by which data are de-identified substantially affects how flexibly and responsively analysts can be to policymakers' and institutional leaders' questions. The moment de-identification occurs is the moment at which available data elements become fixed; after de-identification, no new elements can be appended to the available dataset to answer important questions. Moreover, when de-identified or aggregated data are submitted to a higher-level entity, such as occurs with IPEDS when institu-

tions submit data to the federal government, differences in the construction of data elements, as well as in the cohorts for which data are supplied, can be difficult to detect and their impact on the results will be difficult to evaluate. By limiting the data elements available for analysis and by limiting the group of students who can be analyzed, de-identification and cohort definition inherently put boundaries around how the data collection can be used. By contrast, the NSC collects identified data and supplies identified data to institutions and states for their use. Receiving institutions and states can use those data in combination with their own student data to generate insights about the success of their policies and programs. They can also create a de-identified dataset of these data combinations and provide it to contractors for research and evaluation as their needs dictate. But their ability to make these combinations rests on not prematurely de-identifying the data—that is, institutions do not de-identify the data before shipping it to the NSC and the NSC does not de-identify the data it has matched before sending the data back to institutions or states in response to a request.

Such considerations have obvious implications for privacy and data security. How a state data exchange can ensure that both of privacy and security are provided for will be discussed in more specific terms later. But finding the balance between safeguarding privacy and making data useful is crucial. The key questions to finding that balance are who needs access to what data, and for what purpose? Once the answers to those questions are understood and deemed reasonable, the next questions a state data exchange must address are how should the necessary data be provided, and, crucially, at what point does the data need to be de-identified before being made available? For example, consider a state data system that has, or can get access to, individual-level information on students who were in foster care as youth. The state has a legitimate interest in understanding what is most helpful to such students seeking to access and complete a college education (and what hinders them). The state might want to know, for instance, the extent to which financial aid helped them do so. Yet policymakers looking into the success of former foster-care participants need only aggregate data, while the state's governing or coordinating board for higher education can use de-identified data to run richer analyses on how their institutions and financial aid programs serve those students. But foster-care program administrators whose system served those students could use identified data to better enable them to isolate the effective practices or policies they have in place that may have helped propel the successful students forward. Participation in foster care is not a commonly included data element in existing national data systems nor in those under consideration, leaving states with limited options for evaluating that program. A state data exchange built on maps to data can empower states to better use that data, capturing out-of-state postsecondary and employment expe-

riences to more fully account for those students' outcomes, and to do so in a way that restricts access to the level of data appropriate to the need.

The flexibility to tap richer programmatic and other data affords states that participate in a data exchange the ability to address strategic and tactical uses that go beyond aggregate performance metrics. This includes evaluating programs and policies and gathering information that could lead to better alignment between existing (and aspirational) statewide workforce needs and the set of academic and vocational programs available at public institutions collectively throughout the state. Like SLDSs, such uses can provide states a better understanding of how well the education their students receive contributes to postcollegiate success, and to changes in state and institutional policies, practices, or priorities built on that more complete knowledge. In particular, the ability to disaggregate data and track custom cohorts affords flexibility in evaluating performance, diagnosing problems, and targeting interventions. So if a state is unsatisfied with its performance on a broad metric like the employment rate of its graduates, it could use data from a state-based exchange to dig into that metric to identify the kinds of students, programs, or institutions that are dragging the overall score down—or, conversely, boosting it—and use its analysis to spur improvement. It is not difficult to imagine instances where states might use a data exchange's capacity to link their own rich data to discover opportunities to tweak programs for better results. For example, a state might be interested to know how the employment rate varies among former students with different cocurricular experiences (e.g., co-ops or internships, summer bridge programs, service learning and other leadership activities), and use that information to promote strategies that the data show to be associated with better outcomes. Or, given the sweeping demographic changes buffeting the nation's higher education landscape, a state may want to examine the eventual employment outcomes of its first-generation students. Even if that data element was not among those shared in the data exchange, the state could still use data from the state-based exchange to capture employment outcomes from other participating states and connect those outcomes to its own first-generation data element. Moreover, a state's use of a data exchange would not necessarily be limited to the data it holds in its own data system nor to data elements with common statewide definitions. Rather, it may be possible for the state to forge agreements with individual institutions where the institution provides data to the state for an analysis of, for example, how learning outcomes data the institution may have (but the state does not) are related to its alumni's employment and wages.

In terms of policy and program improvement, states could use a state-based exchange to gather a more complete set

of employment outcome data. They could use these data to better understand the extent to which students who received state-funded merit or need-based awards remained in state after graduation, and for how long. Additionally, the state of Texas recently set a strategic goal that its public school graduates' median debt does not exceed 60 percent of median wages.²⁸ This goal puts Texas at the forefront of thoughtful uses of linked education and employment data for strategic planning usage. But having access to data on the students who leave the state would be important to consider, especially if a geographically smaller state with larger populations along the borders sought to copy Texas's example. States could also find strategic uses of the data exchange to create novel metrics like how cumulative debt relates to earnings in order to estimate the repayment burden of former students, an indicator that could be helpful in informing state and institutional finance policy. Fully realized, the MLDE is well positioned for use in a wide array of applications that can meet the disparate needs of numerous stakeholders. **Appendix A** includes a handful of examples of policy- and practice-relevant research and evaluation questions to better illustrate how states can benefit from sharing data with one another through a state data exchange.

Finally, being owned and governed by states, the MLDE would ensure that states' data needs remain at the forefront of its use, where both state legislators and institutional leaders can take advantage of a resource not truncated by state borders to get answers to the most critical policy and practice questions they face. State ownership also means that states know how their data are being used, and how to best ensure reliable results. In other words, data reporting is not viewed as primarily a compliance-oriented requirement with limited utility to the state or institutions. Rather, state data owners must steward the resource, and if the data are actively used by states and institutions, there will be pressure to look after the quality of the data with care.

Key Facets of a State-to-State Data Exchange

Before a state data exchange model can become a widely used and sustainable resource, and a core component in an improved national postsecondary data infrastructure, its approach to addressing several especially key issues must be clarified. Fortunately, the work by WICHE and its state partners has identified many of the most significant of these issues and has made substantial progress in developing solutions, as discussed below.

Privacy protection and data security. A state data exchange is not alone in confronting the need to address each of these challenges,²⁹ as the intense political climate around this topic, and the very real potential threats to data security, will have to be addressed by all proposed solutions involving individual-level data. Such concerns are magnified since in

an optimized data exchange among states, the participating states should be able to gain access to personally identifiable information available only through external administrative data systems (e.g., another state's, a federal data source, or a third-party data source such as the NSC). This principle is grounded in the notion that if states are expected to figure out how to improve students' success, then they need information about how the students they served performed in subsequent educational settings or in the workforce. And, in fact, state education agencies can already access personally identifiable information through the NSC and some, like Washington's Education Research and Data Center and its State Board for Community and Technical Colleges, can even get personally identifiable data on workforce participation within the state. Given the potential for more effective use by institutions, it is unfortunate that institutions themselves generally cannot receive individual-level UI wage records, even in de-identified form, unless they are classified as a state agency. That leaves many public institutions, and all private institutions, unable to analyze their own students' employment outcomes.

To address concerns about both privacy and data security, the MLDE project has focused considerable attention on honing the architecture for the exchange, as well as its governance. As previously described, the federated data model under development takes care to limit the set of personally identifiable data being collected and secures those data offline. It relies on a series of randomly generated identifiers to permit states to request specific enrollment and employment records from one another. The host of the data used for matching identities in a federated state-based exchange must be held to strict standards for data security, such as ensuring that personally identifiable data are encrypted at rest and in transit. Additionally, the data sharing agreements on which the exchange will operate must include very clear limitations on access to and use of data obtained through the process. As an example, in the MLDE, the agreements expressly prohibit participating states from using MLDE-sourced data in any way that applies to any individual student (including making determinations on educational status or eligibility, enforcing any educational or other action, providing academic advice or support, and marketing goods or services), and that any analytical results may be publicly released only in aggregate.

Voluntary participation is a political strength but an analytical limitation. In an environment in which some stakeholders express skepticism of the federal government and its motives related to data collection and use, the fact that participation in the MLDE is a voluntary decision on the part of states can be viewed as a political strength. But that strength is offset by gaps in data availability caused by nonparticipation among states. These gaps are sure to be largest while a state-based exchange is under development, although they would shrink

as more states agree to participate and would eventually disappear—at least for students in the public sector—if all 50 states and the District of Columbia, as well as any of the U.S. territories and freely associated states, come aboard. (WICHE's MLDE project aims to have at least 10 states in the fold by summer 2016, with plans to recruit additional states from that point forward.) It may be that political or other factors prevent some states from joining, but precedents exist whereby all states eventually come aboard once the value of the resource becomes unquestionably clear to key state leaders—WRIS being the most directly relevant. Additionally, gaps exist as a result of the necessity of initially focusing on linking states' existing databases together, which has meant that students who enroll at private sector institutions are mostly—though not entirely—left out. By tapping into data available through the NSC, information for students in those sectors is available for states to query, but only if those individuals are included in the MLDE crosswalk table in the first place by virtue of having attended a public secondary or postsecondary institution in one of the MLDE states. Data on their employment outcomes will be available only if they had at some point also attended a public postsecondary institution, where a Social Security Number (SSN) could be used to link into the UI wage record data in one of the participating states.

The absence of data from all postsecondary institutions is a drawback for using a state data exchange for accountability purposes, especially at the federal level. While a number of states allow state financial aid to flow to students attending private nonprofit and even for-profit institutions, federal interest in how Pell Grant and loan dollars are being spent and the associated payoffs is not likely to be satisfied by the data available through a state data exchange. As with eventual participation by all states, lacking private institutions' data is not necessarily a permanent condition, as there are potential avenues for private sector institutions to participate in a state data exchange—such as by voluntarily partnering with the states in which they are located to submit data on their behalf. Signs suggest that there may be some interest growing to do so. The most obvious avenue is by leveraging existing data in state data systems for those students whose data must be submitted to the state as a condition of eligibility for state financial aid programs, or by making it possible for private sector institutions to submit their data on a voluntary basis. In both cases, there would be no significant change in how the MLDE is being set up. Another way may be for private sector institutions to band together in a statewide or multistate data system for their sector and seek to participate in a state data exchange, although accommodating such an approach would require adjustments in both governance and operation. In any case, nothing would prohibit participating private institutions from sharing their students' SSNs in order for their students' employment outcomes to be included in analyses, which would begin to fill in the most

significant coverage gap. States could encourage participation by private institutions in this way by making it a condition of their continued state financial aid eligibility. States could also consider whether to obtain students' SSNs for use in acquiring labor market outcomes information (to be used in aggregate) through linkages with other data sources they have access to, such as state driver's license data. Of course, it is also possible that the federal government could simply require private sector institutions to submit data to a unit-level data system with links to employment data like MLDE, which could almost instantly address this shortcoming. But making participation compulsory would be politically difficult and could severely limit the appeal of a voluntary system, not to mention also turning the MLDE into a compliance-first-oriented resource that likely undermines its utility in policy and program evaluation and improvement.

Data comparability, quality, and limitations in states' analytical capacity. In addition to the challenges of data coverage, a state data exchange effort will have to resolve technical issues related to how individual-level data are combined across data systems and state lines and how to ensure that data elements are similar enough in definition to be easily analyzed. Participating states need capacity to engage collaboratively on developing solutions to these issues, not to mention the capacity they will need to use the data exchange once it is established.

A fundamental challenge states are encountering involves the quality of matchmaking and data cleanliness. The MLDE project has been compelled to wrestle with this issue more than most because the various states participating in the project have each had to sort out how to optimize the likelihood of records accurately matching one another across different systems, typically K–12 and postsecondary. Because the MLDE aspires to break down data barriers across both sectors and states, it must find a satisfactory solution to the identity resolution challenge that does not rely extensively on manual processes yet is sufficiently accurate for use in informing public policy and institutional practice. But little consensus exists for what constitutes quality in identity resolution across K–12 and postsecondary data systems.³⁰ It is clear, though, that the threshold for acceptability depends on the intended use of the resulting data—a higher threshold would be needed for uses that are aimed at individual students than what might be required for reporting data in aggregate for policymakers.³¹ (See **Sidebar 3** for a brief elaboration of the challenges related to identity resolution.) Other possible unit-record data systems will eventually have to address similar challenges, of course. But the problem is less prevalent in postsecondary data systems that are not trying to connect to K–12 education. This is because SSNs are more commonly collected, in large part because institutions must report tuition payments to the Internal Revenue Service for students

to claim applicable tax credits (though it is no less critical to be able to examine and evaluate the quality of those matches as well).³² While acknowledging the need for high quality in the matching of records in a state data exchange, it is at least equally important that participating states understand how matches are being made and what strengths and weaknesses exist in the process. Therefore, a state-based data exchange must actively engage states in testing the configuration of an identity resolution process so that states can have confidence in the results produced, even when the matchmaking is (inevitably) imperfect. This transparency would allow states to more appropriately interpret results derived from data sourced from another state. Furthermore, the governance process overseeing an exchange must recognize that states will need to periodically review and refine that configuration.

SIDEBAR 3: IDENTITY RESOLUTION

The ability to accurately link records across data systems is a technical challenge at the heart of current efforts to develop SLDSs and a state educational data exchange like the MLDE. Yet, unsurprisingly, this problem is not unique to education and has received a great deal of attention throughout the history of computing.³³ In all that time, a perfect solution has remained elusive. In states' efforts to set up their SLDSs, they have taken various approaches to resolving identities caused by bad or missing data in individual students' records. A big source of the challenge comes from K–12 education agencies that mostly do not collect SSNs, leading these systems to make matches based on names, dates of birth, and other attributes, sometimes including enrollment information. But even when agencies do have an SSN, it is not always accurate.³⁴ No matter what algorithms are used, states have discovered that a “match rate” measure is not as straightforward as it appears. This is because both false positive and false negative matches can occur (meaning that it is possible to achieve a match rate that exceeds 100 percent), not to mention that a match rate measure requires the assumption that there exists a set of empirically exact matches, to which the rate could be compared. Even though states are commonly able to accurately match a large share of the students they serve across systems, the remaining number that requires manual matching can be extremely large, leading to unacceptable burdens to get them right. Furthermore, as time passes, states acquire more or better data, often resulting in the combination of two previously unmatched records now revealed to apply to the same person, or the opposite condition. The quality of the match—and its transparency to analysts—is critical to the level of acceptance SLDSs achieve. However, developers of matching algorithms inevitably face choices that have tradeoffs in the precision of the match. A more lenient approach to combining records will link more of them, but at the cost of a higher likelihood of false positives, while a stricter algorithm will produce more accurate matches among those that do get linked, but will more frequently fail to match records that do in fact belong to the same person.

A second consideration is how comparable data elements are across systems and states, which a state data exchange can address in three ways. First, a state data exchange builds from observations that many of the most meaningful research questions of relevance to policymakers can be addressed with relatively few data elements focused on enrollment, completion, and employment outcomes—supported by the ability to match records across data systems and appropriate data governance.³⁵ Getting states to focus on making sure a few elements are comparable is less burdensome than it would be for a much larger group of elements. Second, a state data exchange can ask participating states to align the data elements they contribute with the Common Education Data Standards (CEDS), wherever possible. WICHE's experience to date has been that the relative straightforwardness of the data elements included in the MLDE has made CEDS relatively easy to adopt. Additionally, however, by engaging widely dispersed stakeholders on what can be adequate for all of them, CEDS has helped the MLDE project avoid squabbles about individual states' preferred definitions.³⁶ Third, robust participatory structures necessary for governing a state data exchange offer states the opportunity to sort out inconsistencies in data element definition (as well as iron out any other issues that arise). That is, the governance structure imposes a high degree of collaboration on participating states, with a standing committee likely to be charged in part with addressing issues of data inconsistency and interpretation.

The success of a state-based data exchange will rest on its ability to make insightful and meaningful analyses possible. To do that, states must have or find adequate capacity to conceptualize and perform those analyses. As it develops, a state data exchange can address this capacity need by finding ways to productively employ analytical talent that exists outside the state agencies, such as at colleges and universities. In developing their SLDSs, some states, such as Utah, have intentionally partnered with their universities to ensure that the analytical capacity exists.

A second way to address the capacity challenge will be to work on streamlining processing for data requests through two mechanisms. First, queries to the state data exchange can be restricted, at least initially, to a designated window of time that is aligned with the legislative calendar, under which, with few exceptions, state legislatures are only in session during the first several months of the year. That way, states can be thoughtful and prepared to make requests and respond to them at a time when their agency schedules are less hectic. (At its partner states' recommendation, WICHE anticipates that MLDE states would plan ahead for the analyses they expect to need for the upcoming legislative session. State analysts are likely quite capable of anticipating the questions they will be asked based on what they have faced in previous years.)

Additionally, states may be able to save time and energy by working together through a state data exchange to access other potentially valuable data sources like those listed in **Table 1**. Each of these external, non-state-based data sources has its own rules for accessing data. Keeping track of those differences is a challenge both for states and for the entities administering each data source, which have to respond individually to each state’s request. A state data exchange might offer states a way to broker standardized and streamlined procedures for gaining access to those sources. For example, the exchange might do this by aggregating all states’ data requests in a single query to the data source with sufficient information for results to be distributed appropriately, or by establishing the state data exchange (itself or its contractor) as a contractor to the data source for the fulfillment of states’ data requests. Given its focus has by necessity been on connecting states’ data systems with one another, building connections to these external data sources remains largely aspirational at this point, apart from more advanced discussions with the NSC. WICHE and its partner states anticipate that the federated architecture being built for the MLDE may be sufficiently flexible for these connections to be technically feasible, if each of them is considered to be a “node” in the data transmission process. Such connections would require a different data sharing agreement specifying each external data source’s limits on and procedures related to data access and use, as well as what level of data the states or their contractors would be able to obtain. Even if these sources would be willing to provide only aggregate data, states would benefit from filling in the remaining holes from the linkages the MLDE was able to provide them, like earnings from those exclusively self-employed or employed by the federal government.

TABLE 1. VALUABLE EXTERNAL, NONSTATE DATA SOURCES THAT COULD POTENTIALLY BE TAPPED BY AN MLDE

Data Source	Description of Contents
National Student Clearinghouse	Student enrollment and credential records
Federal Employment Data Exchange System (FEDES)	Nonmilitary employees of the federal government
Veterans Administration	Status as a veteran or a dependent
Department of Defense	Status as an active duty member of the armed forces; branch of the armed services; wages for active duty personnel
Census LEHD	Wages for all 50 states and DC from states’ UI files, plus self-employment and federal employment from tax data
Social Security Administration	Wages for all workers nationally from tax data

Additionally, one of the many challenges state agencies are facing as more data linkages are being made comes from having to examine and evaluate the way in which gaps in available data may bias the results. Filling in known sources of those gaps potentially helps reduce the analytical burden on state agencies that are being questioned about results derived from such linkages. The most obvious such gap is out-of-state employment and, as uncovered in WICHE’s initial pilot, the four participating states were able to fill in as much as 22 percent of the missing data on employment outcomes one year after completion by forging linkages with one another.³⁷ Because more individuals find employment in other states as time passes, we can be confident that more data will go missing. Therefore, capturing that mobility in the datasets that analysts use is the most straightforward way not only to produce the best results, but also to address concerns about the potential presence of bias in those results that arise from missing data.

Governance and sustainability. As previously discussed, state ownership is a significant advantage to the state data exchange model. How to establish a body vested with the authority to govern and manage a state data exchange deserves careful attention and planning. Throughout its efforts to develop the MLDE, WICHE has by necessity played the role of decision maker in setting the general direction of the project and for issues such as how to invest time and money on contractors and how to prioritize analyses. However, when it comes to how states can and will exchange data, WICHE is serving as a broker of agreements since it holds no authority over any of the data. From the outset of the MLDE project, it has been clear that any sustainable state data exchange would need its own governance independent of, or at least with authority over, WICHE (or any third-party contractor hired to handle the exchange’s day-to-day operations). This is partly because, as a regional organization, WICHE is not well suited to permanently govern a resource that transcends the West. As an incubator of a state data exchange, it may be an advantage that WICHE is ill-suited for long-term governance of the MLDE because WICHE is compelled to lead the project by giving the participating states as much authority over decision-making as it can without surrendering forward progress in developing the MLDE. It has also been a necessity that WICHE has invested considerable time and energy in working out what the appropriate structure will be. It is too early yet to know many details. But to be effective, it will be essential for both states and sectors to have voice and representation in the governance structure that emerges, although it is infeasible that each agency in each state can have its own separate representative seated at the governing table.

One possible option for MLDE governance and management would be to transfer those responsibilities to an existing organization with sufficient representation for all states already seated in its own governance structure. Organizations such as the Education Commission of the States, the State Higher Education Executive Officers national association, and even NSC come to mind as possibilities, though none have a mission that covers nor representation that includes the labor agencies that hold the workforce data.

A likelier option would be to create a new 501(c)(3) organization around which robust governance could be built. Early thinking on this would suggest a governance board of nine to 11 officers, serving staggered terms, who have appropriate and substantial responsibilities over relevant data in their respective states—the leader of an SLDS or a state higher education executive officer, for example. Expectations are that, at a minimum, one seat would be set aside for each of the sectors that have contributed data to the exchange (presumably one for K–12 education, one for postsecondary education, and one for workforce/labor market information). That body would be supported by at least two advisory committees comprising appointments by the governance board. One committee would focus on technical and technology-related issues, such as data transmission protocols, and advise on matters related to privacy and data security. A second committee would address issues related to appropriate uses of multistate data along with procedures for the onboarding of new states and for providing access to data and disseminating findings publicly. Governance could be further supported through the presence of individual work groups to look after concerns appropriate to their sectors. Reflecting the broad goals of the MLDE project, WICHE’s initial states are ambivalent about these sector-based groups, acknowledging their potential utility while also unsure whether such groups might unintentionally perpetuate any silos that have occasionally made it challenging for sectors to work together seamlessly on data projects.

A second significant challenge is how the MLDE will be supported financially over time. WICHE anticipates that, so long as no more than 10 states are able to participate in the initial rollout, it has sufficient funds to support the development costs of the architectural model previously described. Given the expectations that states will own the MLDE through one of the governance arrangements described above, states are likely to act as the major funding source for operations moving forward. Here again, the MLDE project is considering approaches to a long-term business model that will work, whether states pay a subscription, pay on a per-use basis, or use some combination of the two. States are not flush with funds to pay for a data resource in which they do not find value. But with so many expansive demands for good information about the payoff to higher education for students and

society, as well as the need to make strategic decisions about how to align educational programming with economic conditions, state engagement in the MLDE has been consistently robust. This suggests there is widespread agreement about the anticipated utility for states in combining their data in a manner that is flexible to the needs they face.

For ongoing maintenance and support costs after foundation support runs out, WICHE anticipates that costs will include the following categories:

- ▶ An entity, likely a contractor, tasked with overseeing MLDE operations and associated staff (This contractor’s responsibilities will include convening the governance body; managing the intake process for the annual updates to the crosswalk table; working with states to help them respond to data requests from their peers; ensuring strict adherence to data security standards, privacy protections, and allowable use; marketing the MLDE to leaders in participating states and to new participants; managing the onboarding of new states or new sectors in states; providing outreach and service to additional, related data sources, like the Federal Employment Data Exchange System [FEDES]; and initiating requests for and preparing standardized reports and analysis, as well as disseminating those results.)
- ▶ Licensing and possible subcontracting costs for software and hardware associated with the identity resolution process and data storage
- ▶ Meeting expenses, travel (if necessary), and preparation for governance-related meetings, including face-to-face meetings for the governing board and virtual conferences for advisory committees and sector committees
- ▶ Other costs, such as those that may become necessary to remain compliant with a changing legal and regulatory framework around data security, privacy, or use

The combination of these costs will depend on the number and size of states participating in the MLDE, but WICHE expects that economies of scale will apply and that the total annual costs to cover these four categories are likely to be between \$500,000 and \$600,000 for the initial 10 states. These costs would rise as additional states signed on to participate, but the increases are expected to be marginal since the costs of data storage and transmission are modest. Adding a large state like California would be relatively costlier than adding a small state, but it would likely bring substantial additional benefit to the data exchange for existing participants. This estimate is based on the draft contract currently being negotiated with a contractor, and further informed by discussions with information technology project leaders in participating states and elsewhere. Additionally, preliminary estimates provided by NSC to integrate its StudentTracker service—which would return data on student enrollments in private institutions and public institutions in other states

not participating in the MLDE in response to a query from the MLDE—would cost \$70,000 annually.

These are not the only costs associated with the MLDE, however. States that participate will also face costs for “plugging” their systems into the MLDE. It is possible to set up the MLDE data pathways to be largely automated, but the actual costs of doing so depend on how each state has set up both its system (e.g., what platforms it uses, how it staffs data sharing activities) and the way it conducts data governance (i.e., Can the SLDS—or individual sectors’ data systems if the SLDS is unable or unwilling—be used to connect to the MLDE technically and legally under the state’s specific laws and related interpretations? Does the state use a federated data model or a warehoused one? How many connections—to the education agencies and the workforce agencies—are needed?). Answers to these questions may make plugging in for any given state relatively cheap or costly, but in general WICHE does not anticipate technology costs to be a major barrier for states since states typically have experience linking to other systems, which means they have the capacity to transfer data securely, and since neither the data elements themselves nor the anticipated queries are likely to be particularly complex. Once the connection is in place, ongoing costs are expected to be manageable so long as the state does not make substantial changes to either its technology or its governance.

The sustainability cost for states is harder to capture, and it comes in the form of in-kind contributions from staff who will be involved in preparing data for the crosswalk table or for queries, responding to other states’ queries, evaluating data quality, and producing reports. WICHE has already benefited substantially from considerable in-kind contributions by representatives of state agencies who are developing the technical architecture, governance, and use cases. Those contributions will have to continue as the project moves forward, especially as the MLDE works to evaluate the identity resolution configuration being put into place and as states are asked to build a connection to the MLDE’s data pathways. It is worth noting that the significant contributions already made are a clear indication of how valuable the participating states view the MLDE’s potential to inform policy and practice.

Beyond the initial development efforts as states start to use the MLDE, states are reporting that they do not expect large commitments of time and energy to providing the necessary data to the MLDE, or to responding to queries, if the states can agree to restrict those requests to certain times during the year. Current participating state labor agencies, which have the right to request reimbursement for UI wage-record matching, have plans to do so on a quid pro quo basis. Therefore, apart from time commitments, the MLDE is not expecting to face substantial payments for routine request fulfillment, though certain specialized requests may incur a cost,

such as might be the case if an external researcher was given approval to request a customized dataset. Time, however, is a valuable commodity. Initially, MLDE use will almost certainly be more time-consuming for states that opt to use it as they try to evaluate data quality and coverage for the analyses they want to produce. It is difficult to estimate how long it will take for states to get up to speed, given how varied states are in how they have designed, governed, and staffed their education data systems. Experience from the original four states is not likely to help because those states were blazing trails during the initial pilot phase. As more states come on board (and fill gaps in data coverage) and data quality and familiarity improves, WICHE anticipates that those costs would decrease. In particular, if states are already conducting analyses to which MLDE-sourced data are just being added, then the analytical costs are likely to be marginal and the MLDE data may even pay for themselves if analysts are able to avoid some of the effort required to evaluate the impact missing data may have on the results.

Recommendations

WICHE’s experience in leading the MLDE has resulted in a number of insights about how federal and state policymakers can help ensure that relevant data are available to the right stakeholders to most effectively inform public policies and institutional practices, which will lead to more student success and provide better consumer information. What is clear is that only if the federal government mandates institutions to participate in a state data exchange would it be adequate to fulfill the most commonly expressed requirements the federal government has outlined concerning accountability for financial aid and consumer information.³⁸ Yet a state data exchange can contribute in uniquely powerful ways to improving the nation’s postsecondary data, especially in helping states marshal evidence in support of policy and program improvement and in establishing and monitoring their institutions’ performance against strategic goals. To fully realize the potential of a state data exchange, we make the following recommendations, some of which are also relevant to other proposed efforts to improve the nation’s postsecondary data infrastructure.

Recommendations for Federal Policy Makers

1. **Elevate the priority the federal government places on cross-state data partnerships.** Through the SLDS and WDQI programs, the federal government can take a lot of credit for the advances states have made in assembling and beginning to use unit-level data both to better serve students individually and to better understand the ways in which education and workforce training programs are contributing to workforce and economic development locally, at the state level, and nationally. Both programs have included provisions that encourage states to find ways to link their data together as they prepare funding

applications. Nonetheless, states have underemphasized cross-state data exchanges in their responses to those applications. The reasons for this lack of emphasis are relatively clear and not altogether illegitimate: The federal government has supplied much of the funding impetus for state data systems development but has not prioritized cross-state data products. Consequently, states shy away from working with one another, because such collaboration could be muddled by having to figure out who gets how much money and for what and by accountability to the federal government for performance on a grant that would be more diffuse, and the SLDS project's Requests for Applications have not clarified matters on those occasions when they have included any indication that cross-state activities would be supported. Failure to place a strong emphasis on such partnerships sharply restricts how much value state data systems can provide about postsecondary outcomes, especially at the postsecondary level. The federal government could send a message about the importance of cross-state data linkages by showcasing insights gleaned from multiple states' combined data. Even better, it could require future SLDS and WDQI grant applications to include cross-state data partnerships. Taking such a step would come at very little cost to the federal government, but it would add considerably to the utility of the state data systems development projects it is funding and go a long way toward putting actionable and irrefutable data on educational outcomes in states' hands. It is not hard to imagine that analyses built from better and more complete data at the state level would inform federal policy-making, as evidence of such positive results already exists. The federal government could also require states and institutions to participate in a state data exchange effort through reauthorization of the Higher Education Act or some other legislation, or otherwise create strong incentives for states to do so as a requirement of a competitive grant program like Race to the Top. Certainly, a federal requirement would rapidly bring more institutions and states into the fold than would less directive approaches, but it could prove counterproductive if the states and institutions that are compelled to participate are not fully committed to the effort or if the politics around a federal requirement become a diversion. Recent experience, including MLDE but also WRIS, has shown that a voluntary data sharing effort among states can develop its own momentum toward uniform participation, even in the absence of a federal requirement.

2. Fund analytical work that demonstrates how mobile our society has become and the importance of accounting for such mobility in examining educational outcomes. The federal government has assembled impressive data resources for better understanding the impact of individual mobility on labor markets in the American Community Survey and the Census LEHD program. As the effort to examine

employment outcomes for recent college graduates builds momentum, the federal government should fund and highlight careful analytical work that demonstrates how mobile our society has become and how that mobility must be accounted for in drawing conclusions about the value of various kinds of postsecondary credentials. These observations can help state policymakers have more informed policy debates about how their economic development aims are being met through their own educational investments and talent importation. With mobility data comes the capability to conduct truly longitudinal analyses, and both federal and state policymakers would benefit from better information about career trajectories. This information would provide important details about how Americans are making their way in and out of education and training programs to gain the skills they need to be productive in the labor market, as well as in and out of local or regional markets. These insights could help reveal how individuals with different academic and vocational backgrounds are finding success, or not, over a longer period than the current focus on short-term earnings permits and with a richer evidence base than snapshots taken at various intervals provide.

3. Make it possible for states to access and use information housed in federal data systems in order to more comprehensively and accurately analyze employment outcomes and evaluate educational policies and practices. States own a great deal of the data they need for this purpose, with these sizeable exceptions: They cannot easily capture employment outcomes of former students who are self-employed, employed by the federal government in a civilian role, or serving in the nation's armed forces. These significant blind spots are likely to have different effects on states depending on the predominate industries in their economies and the presence of military installations or large federal agencies. States have the most straightforward access to data on federal employment through FEDES, but gaining access to information on the self-employed and active duty military personnel has been difficult or impossible. The federal government can work with states collectively to help ensure that all postcollegiate employment experiences are adequately captured in a manner that ensures that states have the tools they need to manage their institutions and provide useful information to consumers.

4. Require submission of additional data elements to UI wage records. The federal government specifies what data states must collect at a minimum for the administration of UI. It could expand the minimum requirement by adding three data elements that would greatly improve the information data systems are able to provide to policymakers and the general public. First, occupation is an essential element if UI wage records are to be effective at evaluating whether a graduate's employment is in a field related

to his or her program of study. The second data element, hours worked, is collected by a number of states, including Washington and Oregon from the MLDE pilot, and would permit analysts to calculate a wage rate more reliably and consistently. This is important because UI wage data in most states include earnings aggregated over a fiscal quarter, which requires analysts to calculate a wage rate on the basis of a series of assumptions that may be imprecise. Typically they do this by excluding individuals with wages below an amount representing a year's worth of employment at minimum wage.³⁹ By contrast, states like Washington and Oregon that have data on hours worked simply put aggregate quarterly earnings over hours worked in the quarter to get an accurate wage rate. Not only is this strategy preferred, it also helps those states address data quality issues if the resulting calculation shows a wage rate below minimum wage. Finally, a third data element would specify the location of an employee's workplace, which would allow analysts to pinpoint where within a state that person works. For large employers with many locations, such as grocery chains, that report UI wage records from a single headquarters location, it is typically difficult to distinguish where employees actually work. Such information would be helpful in understanding how well distributed human capital is in a state, and be especially valuable for local governments and business associations. Mandatory enhancements imposed by the federal government would undoubtedly run into political opposition from quarters antithetical to government data collection activities and from some corners of the business community, especially small businesses. However, there appears to be reasonably widespread support for the kind of information these (and other) additional elements might provide and, in any case, payroll processing firms have reported that they can relatively easily accommodate enhancements.⁴⁰ If the political forces are currently too strong for the federal government to require it, states can make progress on their own, as recommended below.

Recommendations for State Policymakers or Agency Leaders

- 1. Demand information that captures the outcomes of all students by joining a state data exchange like the MLDE and by using existing state data sharing resources.** In exercising governance over public institutions, state policymakers can establish expectations about what kind of performance in employment outcomes institutions and programs should be able to demonstrate (whether or not they elect to connect funding to metrics that would correspond to those outcomes). They also can require that information about outcomes be made available to help prospective students make choices about where to attend and what to study. But if they choose to do so, state policymakers should make sure all available information is on
- 2. Ensure that state data systems include elements needed to produce and use disaggregated results to ensure equity in education and employment outcomes by program and for specific population groups.** It is evident that employment outcomes can vary substantially based on a student's academic or vocational program, and wide disparities in mission and program mix means that looking at those outcomes only at the institutional level may be misleading. Moreover, many states (as well as the nation as a whole) are experiencing a dramatic demographic shift in which traditionally underrepresented populations are now the fastest-growing groups. Failure to make progress in closing equity gaps has implications for how economically competitive our society is likely to be in the future, given how central educational attainment is to that competitiveness. In addition to racial and ethnic minorities, states also need to be able to zero in on outcomes for students who are receiving state (and federal) financial aid or who have academic deficiencies identified at the outset of their post-secondary career, as well as other groups they believe to be worthy of special focus. Information about employment and repayment is particularly needed for students who do not complete a credential, although much of the attention devoted to education and employment data linkages has concentrated on graduates' outcomes. Thus, state policymakers need disaggregated data to ensure that the measures they are using to hold institutions accountable are complete, but also to ensure that they and the institutions themselves have the tools to make policy and program improvements.
- 3. Require additional data elements be submitted as part of employers' UI data submissions.** When it comes to enhancing the utility of wage records data, states do not have to wait for the federal government to require the col-

hand, not simply be satisfied with incomplete information that the MLDE project has already proved to be accessible with the right set of agreements and political will. Such information is critical to understanding how geographic location, program mix, and local economic conditions can influence metrics like employment rate, median wages, and repayment. State policymakers can do so by insisting that their state is engaged in a state data exchange project like WICHE's MLDE in order to help shape its development. They can also demand that the analyses of educational outcomes that they consume have included data from existing resources such as NSC, WRIS2, or Census LEHD. They can also convey to their colleagues in the federal government how essential it is that they have appropriate access to valid and complete information and that, accordingly, states need a means to obtain information available only by connecting their own data to federal data resources like those described earlier in this paper.

lection of occupation, hours worked, or workplace location, as we have recommended above. In fact, Nebraska and Louisiana have asked employers to voluntarily submit occupation data as part of their UI submissions, and each plan to require that element's inclusion in the future. Their example, along with Alaska's long history of mandating that field, shows that states can move forward without a federal requirement.

4. Ensure that data elements necessary to connect education and employment are available for use.

States are making progress in linking education and employment data, even if not all have yet done so or are making widespread use of those linkages.⁴¹ State policymakers can take two important steps to ensure the most effective use of employment outcomes data, first by recognizing that the key to accessing employment data in the first place is having an SSN on hand. That means that states can resist the understandable temptation to put unnecessary restrictions on the collection of SSNs without jeopardizing data security or individual privacy. Secondly, nowhere is the gap in information on employment outcomes more apparent than for those individuals who do not go to college. State policymakers need data on those students' outcomes as much as they need information on the return on investment for postsecondary programs, but rarely is that information available for recent high school exiters because, without an SSN, those data cannot be linked. Naturally, securing and protecting the SSNs—especially their linkage to other identifying information like names and dates of birth—must be given the highest priority. But the lessons of the MLDE show that it is possible to keep these pieces of information separate and, therefore, more secure.

5. Allow institutions to use employment data from state UI wage records at the unit level.

Generally, state laws consistently prohibit any organizations that are not state agencies from having access to unit-level data, except as contractors to such an agency. But aggregate results alone are limited in how useful they can be to institutions trying to improve policies and practices. To foster the use of employment data for policy and program improvement, states can permit institutions to use unit-record employment data from the UI wage record files, at least in de-identified format. Loosening that restriction would allow institutions to better analyze and adapt what they are doing in their curricula, student support services, or administrative processes that produce the employment outcomes that are revealed in aggregate. Without unit-level data, institutions are unable to use evidence that ties what they are doing to the results they are getting. Institutions trying to improve their aggregate results on a metric like employment rate or

median wages are obliged to make changes unsupported by evidence. State policymakers can address this blind spot either through legislation or by making it clear to the agencies responsible for administering the UI program that institutions may legitimately receive those data for their own students.

Ideally, the many stakeholders involved in delivering postsecondary education will be able to take full advantage of an improved national postsecondary data infrastructure. Policymakers will be able to hold institutions accountable for the outcomes their students achieve (or do not achieve). Prospective students and their families will be able to access good information about the expected payoff for both the institution they choose and the program they are contemplating. And both policymakers and institutions will be able to access data that allow them to shift strategic priorities and to improve policies and practices that lead to better student outcomes. Accomplishing these diverse aims and meeting the objectives of an effective integrated postsecondary data infrastructure likely require multiple solutions that collectively meet the various needs of different stakeholders. States are clearly one such stakeholder, but they cannot fully meet their data needs by operating independently of one another. States have recognized that gap and have been active participants in initiatives to help close it, including WRIS2 and MLDE, proving that state-to-state data linkages have an important role to play in improving the postsecondary data infrastructure. A state data exchange that makes it possible to combine education and employment records is the most flexible option to address states' needs for better understanding how policies and programs lead to desirable outcomes, and for them to act on that information by designing and implementing targeted interventions or adjusting policies for better impact. As the most significant effort to build such a resource currently underway, the MLDE project is building a data model to deliver on that promise and to attract states to agree to share data. As it did in its first phase, it will build from the significant investments in data systems already made by states and the federal government to provide value in holding educational institutions (and workforce training programs) accountable, provide better and more complete information to consumers, and equip states with a tool for monitoring performance and promoting improvement in their own policies. Already an effort that holds tremendous promise as among the most flexible data initiatives being pursued, its value to states, institutions, consumers, and the federal government will grow as more states participate in the MLDE.

Endnotes

- 1 Rorison, J., & Voight, M. (2015). *Weighing the options for improving the national postsecondary data infrastructure*. Retrieved from <http://www.ihep.org/research/publications/weighing-options-improving-national-postsecondary-data-infrastructure>; Voight, M., Long, A. A., Huelsman, M., & Engle, J. (2014). *Mapping the postsecondary data domain: Problems and possibilities*. Retrieved from <http://www.ihep.org/research/publications/mapping-postsecondary-data-domain-problems-and-possibilities>
- 2 A fuller discussion of the background of the MLDE and the lessons learned during the initial pilot project undertaken between 2010 and 2014 can be found in Prescott, B. T. (2014). *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.
- 3 Prescott (2014). Had data from additional states—most notably California—been included, or had the pilot been situated in a part of the nation where cross-border mobility is more a daily routine than it is in the vast, rural West, those figures would surely be much higher.
- 4 Bransberger, P. (2014). *A glimpse beyond state lines: Student outcomes from WICHE's multistate longitudinal data exchange pilot project*. Boulder, CO: Western Interstate Commission for Higher Education.
- 5 Information on how many in-migrants a state was able to attract was provided to states in aggregate by WICHE, since individual states were not able to receive data on students they did not educate.
- 6 Federal datasets like *Baccalaureate and Beyond* can supply information only about how far from their institution a graduate moves—not whether they crossed a state border in the process—only for baccalaureate degree earners, and without making it possible to isolate patterns applicable to individual states. Census data can provide state-specific data on patterns of mobility for individuals with different attainment levels, but without the capacity to zero in on recent graduates or specific academic or vocational programs, and without the capacity to tie those data to the educational experiences of movers.
- 7 Armstrong & Zaback (2016). *Assessing and improving state postsecondary data systems*. Boulder, CO: State Higher Education Executive Officers.
- 8 Kelderman, E. (2013, September 9). Texas' technical colleges are banking on student earnings. *The Chronicle of Higher Education*. Retrieved from <http://chronicle.com/article/In-Texas-TechnicalColleges/141467/>; Florida Board of Governors. (2014). *Performance funding model overview—May 2014*. Retrieved from http://www.flbog.edu/documents_meetings/0185_0764_5516_563%20BUD%20Performance%20Funding%2010%20Metric%20Model%20Condensed%20Version%20for%20Board%20mtg%201-13.pdf
- 9 Examples include: College Measures, which publicizes earnings information drawn from linked education and employment data within several states; Salary Surfer, a web-based tool that shows earnings for recent graduates from California community colleges; and SeekUT, a similar tool marketed by the University of Texas System that shows employment outcomes for its campuses; among others.
- 10 U.S. Senate Committee on Health, Education, Labor, and Pensions. (2015). *Federal postsecondary data transparency and consumer information: Concepts and proposals*. Retrieved from http://www.help.senate.gov/imo/media/Consumer_Information.pdf
- 11 For instance, students completing high school are increasingly likely to attend an out-of-state institution, a pattern that may be fueled by public institutions' recruitment efforts in the face of declining demographics and heightened pressure to fund their operations through tuition dollars. See Prescott, B. T. (2013). *Demography as destiny: Policy considerations in enrollment management*. Boulder, CO: Western Interstate Commission for Higher Education. Cross-border movement does not stop during college, either. See Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. Washington, DC: U.S. Department of Education; Shapiro, D., Dundar, A., Wakhungu, P. K., Yuan, X., & Harrell, A. (2015). *Transfer and mobility: A national view of student movement in postsecondary institutions, fall 2008 cohort* (Signature Report No. 9). Herndon, VA: National Student Clearinghouse Research Center. Retrieved from <https://nscresearchcenter.org/signaturereport9/>. These gaps in state data systems are particularly problematic for measuring employment outcomes, given research that demonstrates how individuals with more education are likelier to be mobile in search of employment opportunities or better wages, whether they do so by relocating or simply commuting. See Wozniak, A. (2010). Are college graduates more responsive to distant labor market opportunities? *Journal of Human Resources*, 45(4), 944–970; Molloy, R., Smith, C. L., & Wozniak, A. (2011). Internal migration in the United States. *Journal of Economic Perspectives*, 25(3), 173–196; Colorado State Demographer. (2015). *What paradox? Educational attainment in Colorado by place of birth*. Denver, CO: State Demography Office. Retrieved from <https://drive.google.com/file/d/0B2oqdpZKJqK7RDl0ZG45V2JmNmc/view?pli=1>. The evidence is also clear that the academic program an individual completes influences wage outcomes. See Carnevale, A. P., Rose, S. J., & Hanson, A. R. (2012). *Certificates: Gateway to gainful employment and college degrees*. Washington, DC: Georgetown University Center on Education and the Workforce.
- 12 Stevens, D., & Zhang, T. (2014). *Toward a business case for sustained investment in state longitudinal integrated data systems*. Baltimore, MD: The Jacob France Institute.
- 13 Namely, those elements are names, dates of birth, and 4- or 9-digit Social Security Numbers—which will be hashed and encrypted to prevent compromising them.
- 14 Space limitations preclude a fuller description here of how the architecture will work; more information can be obtained on WICHE's MLDE project webpage at www.wiche.edu/mlde or by contacting project staff.
- 15 Ewell, P., & L'Orange, H. P. (2009). *The ideal state postsecondary data system: 15 essential characteristics and required functionality*. Retrieved from State Higher Education Executive Officers website: http://sheeo.org/sites/default/files/publications/ideal_data_system.pdf; Prescott, B. T., & Ewell, P. (2009). *A framework for a multi-state human capital development data system*. Retrieved from Western Interstate Commission for Higher Education website: <http://www.wiche.edu/info/publications/FrameworkForAMultistateHumanCapitalDevelopmentDataSystem.pdf>
- 16 Zinn, R. (2016). *Classroom to career: Leveraging employment data to measure labor market outcomes*. Washington D.C.: Workforce Data Quality Campaign.
- 17 Stevens & Zhang (2014).
- 18 According to Armstrong and Whitfield (forthcoming), 18 state higher education executive offices currently collect data from private, nonprofit institutions and 30 collect data from proprietary institutions, although not all of those collections cover all students enrolled at those institutions. Armstrong, J., & Whitfield, C. E. (forthcoming). *Strong foundations*. Boulder, CO: State Higher Education Executive Officers.
- 19 Voight et al. (2014); Engle, J. (2016). *Answering the call: Institutions and states lead the way toward better measures of postsecondary performance*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from <http://postsecondary.gatesfoundation.org/wp-content/uploads/2016/02/AnsweringtheCall.pdf>
- 20 Dundar, A., & Shapiro, D. (2016). *The National Student Clearinghouse as an integral part of the national postsecondary data infrastructure*. Herndon, VA: National Student Clearinghouse Research Center.
- 21 The NSC itself may be a viable alternative to a state data exchange if it could use its data to make connections to a source of employment information, such as the states collectively or by collaborating with a source in the federal government, such as the Social Security Administration. Several obstacles stand in the way of the NSC going it alone, however, including the following: State education agencies (among others) expect transparency in the record-matching algorithm, so they can independently verify and understand its quality. NSC's use of its data for any purpose is governed by agreements it has with postsecondary institutions. This means that even when the NSC has Social Security Numbers (SSN) in its records, it cannot use them to make matches to external data sources (like UI wage records or Social Security Administration data) without permission from provided by institutions. Nor can it unilaterally name institutions in any analyses it releases publicly, which sharply limits the use of NSC data for institutional accountability or consumer information. Should the NSC seek to use UI wage record data for its employment information, it would need to be named a contractor to all UI wage shops in order to do so. Whether state UI administrators would be comfortable with the NSC in this role is an open question, but in any case it is not clear whether the task of establishing those agreements would result in a cost savings to the effort over a state-based approach. The NSC operates a warehouse model for its own collection; adding SSNs and wage data from state sources would likely be a nonstarter for many states. It is unclear if the NSC would be interested in setting up and managing a duplicate collection via a federated model as the MLDE is developing, but doing so would likely be costly.
- 22 Zinn (2016).
- 23 Certain state agencies can get WRIS/WRIS2 data at the unit level, and in some states those agencies also have responsibility for overseeing the community college sector. But the WRIS/WRIS2 data sharing agreements sharply limit the uses to which the data can be put.
- 24 State University System of Florida Board of Governors (n.d.) *Performance funding model overview*. Retrieved from http://www.flbog.edu/about/budget/docs/performance_funding/Overview-Doc-Performance-Funding-10-Metric-Model-Condensed-Version.pdf.
- 25 Personal communication with Jason Jones, Florida Board of Governors, August 17, 2015.
- 26 Schneider, M. (2014). *Measuring the economic success of college graduates: Lessons from the field*. Washington, DC: American Institutes for Research; Carnevale, Rose, & Hanson (2012).
- 27 Prescott (2014).
- 28 Texas Higher Education Coordinating Board. (2015). *60x30TX (Texas higher education strategic plan, 2015–2030)*. Retrieved from <http://www.theccb.state.tx.us/reports/PDF/6862.PDF?CFID=38275436&CFTOKEN=21734195>
- 29 The authors recognize that data security and privacy protection are not equivalent. See Joanna Lyn Grama's paper, "Understanding Information Security and Privacy in Postsecondary Education Data Systems," for a clear discussion of these topics.
- 30 SLDS Topical Webinar Summary. (2015, September 16). *The match rate dilemma*. Retrieved from http://nces.ed.gov/programs/slids/webinars.asp#MW_Sep_2015_01; SLDS Topical Webinar Summary (2015, November 23). *Processes for handling multiple IDs to ensure data quality*. Retrieved from http://nces.ed.gov/programs/slids/webinars.asp#EC_DEC_2015_02.
- 31 Dusetzina S. B., Tyree, S., Meyer A. M., Meyer, A., Green, L., & Carpenter, W.R. (2014, September 4). *Linking data for health services research: A framework and instructional guide*. Rockville, MD: Agency for Healthcare Research and Quality; (2014, September 4). *An overview of record linkage methods*. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK253312/#>; Schumacher, S. (2007). Probabilistic versus deterministic data matching: Making an accurate decision. *Information Management Special Reports*. Retrieved from https://www.healthit.gov/archive/archive_files/FACA%20Hearings/2010/2010-12-09%20Patient%20Linking/Probabilistic%20Versus%20Deterministic%20Data%20Matching.pdf

- 32 The National Student Clearinghouse claims a match rate in the high 90 percent range. Although it has not made its algorithm for making matches public, it does include a process for manually reviewing so-called “near matches.” The College Scorecard relies on SSNs and can be considered reliable because it only includes students with financial aid; extending its analysis to nonfinancial aid applicants and recipients would necessitate a review of the quality and coverage of the SSNs obtained.
- 33 Dusetzina, S. B. et al (2014). Linking data for health services research: A framework and instructional guide. Retrieved from the Agency for Healthcare Research and Quality website: <https://www.ncbi.nlm.nih.gov/books/NBK253313/>.
- 34 Sabel, J. (2014, May 21). *Identity matching: SSNs are not enough!* Presentation at ERDC ARRA SLDS Conference (Olympia, WA).
- 35 Ewell & L’Orange (2009); Prescott & Ewell (2009).
- 36 CEDS in the Field. (n.d.). Retrieved from <https://ceds.ed.gov/pdf/ceds-in-the-field-wiche.pdf>
- 37 Prescott (2014).
- 38 U.S. Senate Committee on Health, Education, Labor, and Pensions (2015).
- 39 See, for example, Colorado’s methodology, which is outlined at EdPays – Education Pays for Colorado website <http://co.edpays.org/>.
- 40 Workforce Information Council. (2014). *Enhancing unemployment insurance wage records: Potential benefits, barriers, and opportunities*. Retrieved from <http://www.workforcedqc.org/sites/default/files/images/Wage%20Record%20Study%20Year%201%20Report-Finalpdf.pdf>
- 41 Armstrong & Whitfield (forthcoming).

APPENDIX A: STATE DATA EXCHANGE USAGE POSSIBILITIES

Use Category	Topic Area	Broad Research/Evaluation Questions	Specific Questions
Accountability	Workforce Innovation and Opportunity Act reporting	What is the performance of WIOA training programs as measured by client outcomes?	How many education program participants earned postsecondary credentials within a year of leaving their program? How many education program participants are employed in MLDE states?
	Financial aid	How well are aid recipients performing in the workforce and contributing to identified workforce needs?	Are aided students retained in our state's labor force after completion? Are students with workforce-contingent grants (i.e., teacher training) finding employment in the state, and how long are they staying in those jobs?
	Institutional performance	Are institutions adequately preparing students for workforce success?	Are programs offered by institutions at least providing a living wage for those who graduate, while also not leaving those who do not with unsustainable debt? Are graduates of some programs clearly outperforming others?
Consumer Information	Return on investment	What is the individual return on investment for those who complete credentials versus those who do not complete?	What are the earnings for those who complete credential X? How do earnings of completers who stay in state compare with those who find employment in other states? How do these earnings compare with those who leave the credential X program short of completion?
Strategic Planning and Monitoring	Affordability	How does students' accumulated debt relate to their earnings trajectory?	What is the median debt-to-earnings ratio for former students with debt in the first year after completing a credential? What is the median debt-to-earnings ratio over the first five to 10 years after completion for former students with debt? How does that differ for students who do not complete a credential?
	Longitudinal time series of human capital development and deployment	How well has education prepared individuals for adapting to rapidly changing economic climates?	How likely and how soon are individuals who earn credential X returning to education programs? How adaptable are completers of credential X as measured by industry of employment and earnings? What is the earnings trajectory for individuals with credential X? How mobile across states and employers (i.e., stability of employment) are they, and what might that convey about educational quality?
	Balance of trade in educated talent	What are the in-migration, out-migration, and net-migration of recent education program completers?	How many individuals who earned credential X within the past five years left our state for another state? How many individuals who earned credential X within the past five years in another state have come to our state to find employment?
Policy and Program Improvement	Student success	What impact does remediation have on expected labor market outcomes? Do our state policies support success in college and beyond for those students identified as in need of remediation?	How well are students who begin their postsecondary experiences with a remedial course finding work, and how do their wages vary, if at all, based on how remedial experiences are provided?
	Curriculum alignment	How well are our graduates prepared to succeed in key industries?	How stable is employment in target industries among our graduates, and how quickly do their wages increase?
	Evaluation of work-related experiences	How well is our programming in work-related experiences, including such things as co-op or internship experiences and involvement in professional societies, working?	Is there a difference in employment outcomes or subsequent enrollment for those graduates who engaged in intentional professional programming? Do some professional programming experiences produce better outcomes than others?
	Tuition-setting and admissions	Are former students who arrived as nonresidents adding to the educational talent of our state's labor market?	How many and what share of nonresidents are finding employment post enrollment in their destination state?

Envisioning the National Postsecondary Data Infrastructure in the 21st Century is a project of the Institute for Higher Education Policy and is supported by the Bill & Melinda Gates Foundation.

