About the Authors

Iris Palmer is a senior policy analyst with New America's Education Policy Program. She provides research and analysis on state policies related to higher education including performance based funding, state student financial aid, and state data systems. Palmer previously worked at the National Governors Association on postsecondary issues. There she helped states strengthen the connection between higher education and the workforce, support competency based systems, use data from effectiveness and efficiency metrics and improve licensure for veterans. Prior to joining NGA, she worked at HCM Strategists on the Lumina Foundation's initiative to develop innovative higher education models, including new technologies and competency-based approaches. Before joining HCM Strategists, Palmer worked at the U.S. Department of Education in all of the offices related to higher education: the Office of Vocational and Adult Education, the Office of Postsecondary Education, the Policy Office and the Office of the Undersecretary. Palmer received her undergraduate degree in Political Science from Goucher College and her Masters of Public Policy from George Mason.

Acknowledgments

We would like to thank the Gates Foundation for its generous support of this work. The views expressed in this report are those of its author and do not necessarily represent the views of the Gates Foundation, its officers, or employees.

About New America

New America is committed to renewing American politics, prosperity, and purpose in the Digital Age. We generate big ideas, bridge the gap between technology and policy, and curate broad public conversation. We combine the best of a policy research institute, technology laboratory, public forum, media platform, and a venture capital fund for ideas. We are a distinctive community of thinkers, writers, researchers, technologists, and community activists who believe deeply in the possibility of American renewal.

Find out more at newamerica.org/our-story.

About the Education Policy Program

New America's Education Policy Program uses original research and policy analysis to solve the nation’s critical education problems, serving as a trusted source of objective analysis and innovative ideas for policymakers, educators, and the public at large. We combine a steadfast concern for low-income and historically disadvantaged people with a belief that better information about education can vastly improve both the policies that govern educational institutions and the quality of learning itself. Our work encompasses the full range of educational opportunities, from early learning to primary and secondary education, college, and the workforce.

Our work is made possible through generous grants from the Alliance for Early Success; the Foundation for Child Development; the Bill and Melinda Gates Foundation; the Evelyn and Walter Haas, Jr. Fund; the HeisingSimons Foundation; the William and Flora Hewlett Foundation; the Joyce Foundation; the W.K. Kellogg Foundation; the Kresge Foundation; Lumina Foundation; the McKnight Foundation; the Charles Stewart Mott Foundation; the David and Lucile Packard Foundation; the J.B. & M.K. Pritzker Family Foundation; the Smith Richardson Foundation; the W. Clement and Jessie V. Stone Foundation; and the Berkshire Taconic Community Foundation.

Find out more at newamerica.org/education-policy.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>How Would a State-based Federal Data System Work?</td>
<td>4</td>
</tr>
<tr>
<td>What Are the Potential Benefits of a State-based Federal Data Approach?</td>
<td>6</td>
</tr>
<tr>
<td>What Are the Challenges to the State-based Data System Model?</td>
<td>7</td>
</tr>
<tr>
<td>Notes</td>
<td>11</td>
</tr>
</tbody>
</table>
One Thursday in September 2012, the House Subcommittee on Higher Education and Workforce Training held a hearing about the sorry state of higher education data. During the hearing, Subcommittee Chairwoman Virginia Foxx (R-N.C.) expressed frustration saying, “We have so much data and we seem to know so little. What a tragedy for all the money that we’re spending in this country.”

Students, families, and policymakers, like Rep. Foxx, want answers to many questions about the performance and outcomes of the higher education system. While the U.S. Department of Education has made a significant step towards transparency with its recent release of new data on student outcomes, many basic questions still cannot be answered at a national level. Answers remain elusive to basic questions like:

- How many non-traditional students attend college and do they successfully complete credentials?
- What happens to students who do not graduate? Are they transferring to other colleges and earning degrees, or are they dropping out altogether?
- How much debt are students accumulating in college, and can they repay their loans?
- Are students obtaining employment in their field after college, and if so, what do they earn?

A few years ago, even many states couldn’t answer these pressing questions about their own public higher education systems, which consume 10 percent of total state spending. In the mid-2000s, the U.S. Department of Education proposed creating a federal system that could help answer these questions. But in the 2008 Higher Education Opportunity Act, Congress barred the Department from developing such a system. This ban amounted to a college blackout, depriving the public and policymakers of information about which schools are doing the best job serving students. Ironically, the primary author of that ban was Rep. Foxx who would later complain about the lack of higher education data at that 2012 hearing. Student groups ranging from Young Invincibles to the Big 10 Student Association have called for lifting the ban so students and their families have the data they need to make informed decisions about colleges and majors.

In light of the ban, the Education Department had no choice but to turn to the states to help answer key questions about educational outcomes. Since 2006, the Department has provided over half a billion dollars to 47 states to build education data systems for this purpose. At the same time, the Department of Labor has spent $35 million to connect these state data systems with workforce...
data. And states have invested even more of their own dollars. Although many of these dollars went only to the K-12 system, the investments have allowed states to answer previously unanswerable questions. Despite this progress, many questions remain, in large part because most state systems lack information about students who attend private colleges and students who leave the state for graduate school, a job, or to finish their undergraduate degree elsewhere.

North Dakota, for example, created its data system to answer questions like: Are students who go to college in the state staying? Are they working in their field of study? Are there enough of them in the education pipeline to meet employer demand? But one problem the state has encountered is that much of its population is centered close to the Minnesota border. Without exchanging data with Minnesota, state policymakers can’t tell where their graduates are going and how they are performing with any level of detail. This means that students who go to college in North Dakota but then get a job just a few miles away in Minnesota can’t be counted. This also means that if a student transfers to a college across state lines, policymakers don’t know if that individual has dropped out or is still enrolled in college.

Policymakers in the state of Washington also had questions. They wanted to know if the state’s college graduates were getting good jobs and if high school students who were in dual enrollment programs were faring well in college (and beyond). They too were losing some students out of state and weren’t fully satisfied with the answers they were able to generate. Unlike in North Dakota, which had not been able to share data with Minnesota, Washington policymakers had a potential solution.

The Western Interstate Commission on Higher Education (WICHE) had started a pilot project that would exchange state data on two groups of students to answer policy questions across four western states: Hawaii, Idaho, Oregon, and Washington. The project, the Multistate Longitudinal Data Exchange (MLDE), ended up plugging significant holes in each state’s stand-alone information on student mobility and labor market outcomes. In Washington, the data exchange uncovered more than nine percent of individuals not found working in the state that they had previously been unable to track with state data alone. WICHE is now working to add at least an additional six states by the summer of 2016 for a total of 10 participating in the pilot. The MLDE project will never be a replacement for, or function as, an information system to answer federal higher education questions. MLDE is meant only to provide states with additional information they do not currently have access to. But WICHE’s experience administering MLDE can shed light on how such a state-based system could be created to answer questions at the national level.

The idea of “stitching together” state data systems, otherwise known as a state-based federal data system, provides an appealing alternative for policymakers eager to understand both state and national-level college outcomes in a way that bypasses the politics of the 2008 ban on a federal system. Two prominent senators, Ron Wyden (D-Ore.) and Marco Rubio (R-Fla.), among others, took up this charge in 2012 with the introduction of the Student Right to Know Before You Go Act. This first version of the legislation would have been in some ways like the WICHE project, but it would apply to all 50 states. The measure would have had states report their student data to a central location outside of the federal government, where the data would be connected and aggregated for federal use and reported back to the states for institutional and system-wide improvements. This seemed like the best way to get student unit record data while not violating the 2008 ban. But as policymakers considered this approach, it became clear that it might harbor its own challenges including a failure to account for outcomes at private institutions. In the face of these difficulties, along with increased calls for better data by a variety of stakeholders, Sens. Wyden and Rubio ditched the work around, and in 2014, simply called for a repeal of the ban and the creation of a federal student record system.
Although the Student Right to Know Before You Go Act has been introduced in several sessions, it has yet to be voted on, let alone to become law. While support for ending the ban appears to be growing, the politics that led to the ban haven’t shifted enough to repeal it. Currently, all major associations that represent only public colleges support the repeal. A federal student unit record system aggregating existing federal data is simple and complete. But the powerful private college lobby that pushed the ban in 2008 hasn’t budged. Some congressional staffers whose members support the ban have expressed interest in the stitch-the-states work-around. Given the continued ban and the pressing need to answer questions about college outcomes, the potential benefits and drawbacks of the “stitch-the-states” approach must be examined.

HOW WOULD A STATE-BASED FEDERAL DATA SYSTEM WORK?

In a state-based federal data system, state systems would submit the student data necessary to identify individuals to a third party. This third party would then match student enrollment data from one state with outcome information from other states, and use that data in an anonymized to answer national policy questions like those from Chairwoman Foxx. This third party could theoretically be any entity capable of aggregating and analyzing the data.

While this hypothetical national data infrastructure is not the intended purpose of WICHE’s MLDE project, the pilot can shed light on how this might work at the national level. Under WICHE’s project, the states reported their education and unemployment wage records to a national nonprofit specializing in higher education data called the National Student Clearinghouse (NSC). At the NSC, the data were cleaned and matched to other states before being reported back to the states with the additional student information from across state lines. Figure 1, from WICHE’s Beyond Borders publication, shows the complexity of the process they designed for exchanging and matching the data for just four states. We explore the complexity in more detail in the challenges section below, but this figure shows how the data must be exchanged between multiple entities in five steps to secure the additional data from across state lines.

In the next stage of the pilot, the data exchange will be structured differently to improve the protection of personally identifiable information. The states will share identifiable information that
has been encrypted with the vendor. This vendor will use an identity resolution engine to develop and adjust a series of matching rules. Once the algorithm is approved, states will upload the encrypted identifiable elements every year to a central crosswalk. That crosswalk will facilitate the matching and produce a table that will route data requests to the participating states. This complex matching system will further protect student privacy.

**Figure 1 | MLDE Initial Exchange of Data Process/Architecture**

<table>
<thead>
<tr>
<th>K-12 Data</th>
<th>Cohort A: High School Graduates</th>
<th>Vendor (NSC)</th>
<th>Labor/UI Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>HI</td>
<td></td>
<td></td>
<td>HI</td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td>ID</td>
</tr>
<tr>
<td>ID</td>
<td></td>
<td></td>
<td>OR</td>
</tr>
<tr>
<td>WA</td>
<td></td>
<td></td>
<td>WA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Postsecondary Education Data</th>
<th>Cohort B: First-time Undergrads</th>
<th>“Core” Exchange Dataset</th>
<th>NSC Main Collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>HI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Postsecondary Education Data</th>
<th>Hawaii Enhanced Dataset</th>
<th>Idaho Enhanced Dataset</th>
<th>Oregon Enhanced Dataset</th>
<th>Washington Enhanced Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>HI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: “NSC” refers to the National Student Clearinghouse. * Except in Oregon, where data went directly to/from the Oregon Employment Department.

WHAT ARE THE POTENTIAL BENEFITS OF A STATE-BASED FEDERAL DATA APPROACH?

**Uses existing infrastructure.** The federal government has invested $611 million in state longitudinal data systems, and states have invested untold millions more. According to the State Higher Education Executive Officers, “There is a substantial and growing amount of statewide, coordinated multi-sector data sharing across the country, much of which likely has been influenced by the U.S. Department of Education Statewide Longitudinal Data Systems (SLDS) grant program.” A state-based system would capitalize on that investment by building on its foundation. In fact, two grant recipients under this program, Virginia and Nevada, have built their SLDS as federated systems, further demonstrating the potential of this model. But many states still don’t have a high quality SLDS and additional resources would be needed to get these states up to standard.

**Doesn’t require removing the ban on a federal student unit record system.** As mentioned above, a federal student unit record system is currently banned by federal law. While the prospects of removing the ban on creating a federal student unit record system have improved in the last few years, it still faces an uphill political climb. A state-based data system could answer some student and policymaker questions without lifting this ban. WICHE concluded, “As an alternative to federal solutions, WICHE’s MLDE has clear advantages tied to its ability to make data available to states.”

**Isn’t controlled at the federal level.** Some Americans are uncomfortable with the idea of the federal government having access to and being able to use this type of personally identifiable data. Educational outcomes data would be collected, in accordance with existing federal and state laws put in place to protect student privacy. However, concerns about the federal government’s ability to adequately protect this data remain. Given this discomfort, it may be easier to sell the public on linking existing state data systems, rather than creating a scary-sounding federal data system. This could be true even though the federal data system would essentially just be linking existing data sources and transferring data between all 50 states may pose its own risks.

**Could allow states to use outcome data from across state lines for more detailed analysis.** Current discussions of a federal system do not include allowing states to access identifiable information to enhance national data with their
own state analysis. The stitch-the-states method, as piloted by WICHE, does allow states to access identifiable information for their students, regardless of whether these individuals remain in-state, creating a more complete understanding of how well students are served. This allows the states, in turn, to enhance inter-state data with their own state analysis to evaluate, for example, the effectiveness of their state aid programs, remedial redesign, or any number of other state specific interventions.

WHAT ARE THE CHALLENGES TO THE STATE-BASED DATA SYSTEM MODEL?

Lack of information about students who attend private institutions. Over two million college students attend private not-for-profit or for-profit colleges.13 The vast majority of state data systems do not include these students. Seventeen states collect some information about private college students in their states, but the data is typically not complete.14 Excluding private universities would leave out a significant number of students especially at institutions that particularly concern policymakers. In addition, the data don’t capture students who may start at a public and transfer to a private college. Since many students transfer at least once in their college career, not being able to see how students progress if they transfer to a private school would significantly blunt the usefulness of the data.

Current efforts are voluntary. In order to answer national-level questions, you need national-level data. The current MLDE data belong to states to use as they see fit, and the project will remain a voluntary effort. The only way to get the states to participate in the pilot was to provide them value without requiring that the data be made public. If participation in the state-based federal data effort was voluntary, states could participate in some years and not in others leaving large holes in the data over time. Unless participation and publication of the data are both mandatory, this system will not meet the needs of students, federal policymakers, or the public. In order for this system to provide a useful level of transparency at the national level—or the baseline for an accountability system—it would need to be required of all states. In the Wyden-Rubio 1.0 bill for instance, all Title IV eligible institutions were required to participate.

Requires a third party to hold, match, and clean the data. A state-based federal data system cannot exist without a technically savvy, trusted contractor,
who holds, cleans and matches the data. This contractor could either be a private entity or a state agency. In the WICHE pilot, this contractor was the non-profit organization, NSC. NSC is a private nonprofit organization, established more than 20 years ago by lenders to help administer the federally guaranteed loan program. NSC currently has one of the largest student-level data sets in the country and is able to link substantial amounts of data across institutions to achieve a fuller picture of college churn and completion. But it can only do this because it has a highly detailed and national-level student unit record system, stocked with student records that the institutions have voluntarily shared with NSC. The use of any private entity in this manner means the government would have to address concerns about accountability, privacy, and pricing. If the contractor is a private entity, how can students and the government be ensured of the privacy and security of that personally identifiable information? And how can we ensure the needed data elements are collected? Such private entities have raised concern in K-12 data policy. Using a private organization, whether for-profit or not-for-profit, will likely raise similar privacy concerns that could be better addressed by the federal government simply collecting and protecting the data. It will also create a government-sanctioned monopoly with control of that data. Another option is to have an existing state data entity run the system. But this would likely bring up a different set of territorial and political challenges. For instance, one data contact in Maine, when asked if another state could be the broker of a state-based federal data system, said, “I don’t think that would fly...we don’t want another state agency to be the broker. We’d rather have a party that isn’t in the game and is interested in linkage but not data.”

Limited earnings information. While the federal government has the most complete earnings information available through the Internal Revenue Service (IRS) and Social Security Agency (SSA), states have access to only Unemployment Insurance (UI) data. Unfortunately, this data has some significant holes, including the fact that it doesn’t cover federal employees—including military personnel—or the self-employed. While some of these gaps can be partially addressed by combining UI data with other data sources like Federal Employment Data Exchange System (FEDES), data held by the IRS and SSA already capture these populations. For states like Virginia, this is a huge challenge, because more than 172,000 federal employees call the state home. Using existing federal tax data like SSA to follow students as they enter the labor market would paint a richer and more accurate picture of employment outcomes for everyone.

If the contractor is a private entity, how can students and the government be ensured of the privacy and security of that personally identifiable information? And how can we ensure the needed data elements are collected?

Lack of trust and data governance issues both inside and across states. The agencies housing these data can be fragmented and territorial about data within their own states. These issues of trust and cooperation are only amplified across states. Getting these government bureaucracies to play nice and share their data can be an exercise in frustration. The WICHE pilot spent a large amount of time addressing data governance concerns and convincing state agencies to work together. For example, in three of the four states, the education data had to pass from the NSC to the agency in charge of the state longitudinal data system, and then to the agency that houses the labor market data, and then back to the NSC. The extra step of passing the data to the entity housing the labor market data was necessary due to data governance. Similar complications would only be amplified across all 50 states. One data expert from North Dakota told us, “Who can see what and the political side of permissions is the hard part. The technical aspect is easy. Then you have to go an extra step and
work that out between states.” If a state-based data model were to become the national higher education data system, this type of barrier would need to be addressed in all 50 states, a massive undertaking.

**Legal concerns about state and federal privacy laws.** Federal laws, the Family Educational Rights and Privacy Act (FERPA) and the Privacy Act of 1974, protect education and workforce data. The Department of Education’s guidance around FERPA helped WICHE get their pilot off the ground. However, attorneys general in different states have different interpretations of what these laws require, and states have their own privacy laws governing these types of data. While they were able to overcome this barrier, WICHE cited “legal concerns with respect to federal and state privacy protections, while preserving confidentiality protections and ensuring data security” as a major barrier to their work.15 Getting all states to agree on what these laws require would be another major hurdle.

**State data systems are at different levels of development.** In the WICHE pilot, Oregon did not yet have a strong state student level data system, which meant each participating agency in the state (K-12, higher education, and workforce) had to share data individually with the NSC. As a result, NSC had to deal with three more agencies and clean and match three more data sets, essentially creating a longitudinal data system for the state. Like Oregon, there are still several states where these silos between agencies are firmly in place. There are even more states, like California, where there are multiple higher education systems, each with their own data. These states have different levels of cooperation across systems. Each of these systems and agencies reporting to the third party vendor would increase the complexity significantly. Instead of dealing with 50 sets of data, the contractor could end up dealing with well over 50 data sets, each governed by a different contract.

**Substantial investment of time and money.** Getting all of the various stakeholders to agree and share data was expensive and slow in the WICHE pilot. WICHE received a grant in 2010 from the Gates Foundation for $1.5 million to support this initial pilot project with only four states.16 At the same time, many of the state systems also contributed resources in time and effort. The work started in 2010, but states didn’t start turning over data until 2012. And there were no final results until 2014. This would speed up as states learned the process, but it could remain very long and involved. WICHE wrote in one of their reports that “improving flexibility and timeliness is important if a future version of the MLDE is to optimally deliver on its potential to inform public policy makers and institutional leaders.”17 We should compare the price and timeline of creating a state-based data system to the price and timeline for establishing a federal student record system.

---

We should compare the price and timeline of creating a state-based data system to the price and timeline for establishing a federal student record system.

**Technical issues.** There were a slew of technical issues the WICHE pilot ran into that would need to be solved in order to scale a state-based federal data system across the country. Here are some of the issues that the pilot program struggled with:

- It took a lot of time and effort for the four states to develop common data definitions, although the existence of the Common Education Data Standards project did help with this. The Common Education Data Standards project is a national collaborative effort to develop voluntary, common data standards for a key set of education data elements to streamline data exchange. But states are at different places in their engagement with and implementation of the Common Education Data Standards. A state data contact from North Dakota said, “One of the things that would help everyone is if everyone spoke in the same language...defined at the federal level.” But that is not yet a reality.
• There was no clear rule on what level of misidentification of students was inevitable/tolerable.

• Cleaning the data was difficult and time consuming for both the states and the NSC. For example, NSC found a significant number of students reported as first time in the cohort who were, in fact, continuing students.

These types of technical issues will most likely be challenges in any new postsecondary data infrastructure but the MLDE pilot highlighted them in the stitch-the-states model.

There are some advantages—the most obvious being a political advantage—to using a state-based federal student unit record system over a federal student record system. The federal government has invested heavily in state data systems. States such as Virginia and Florida already have rich data systems that, if nationally linked, would be capable of answering many critical questions while eliminating time-consuming federal reporting. What’s more, some states are already linking data across state borders to see how students fare as they move through the region’s educational systems and the workforce. But this approach also has numerous challenges. How would private colleges be addressed? Who would be in charge of matching the data? How would we overcome the multitude of technical challenges on a larger scale? While a state-based data system can be developed, it is far from an ideal replacement for the current higher education reporting system. Instead, creating a single, federal student unit record system could be a simpler and faster way to end the college blackout.
Is Stitching State Data Systems the Solution to the College Blackout?

Notes

1 See College Blackout, Clare McCann and Amy Laitinen, New America Education Policy Program https://www.newamerica.org/downloads/CollegeBlackoutFINAL.pdf.


4 See College should be like buying a car, Keaon Dousti http://t.co/HlRxz1otjF.


This report carries a Creative Commons Attribution 4.0 International license, which permits re-use of New America content when proper attribution is provided. This means you are free to share and adapt New America’s work, or include our content in derivative works, under the following conditions:

- **Attribution.** You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

For the full legal code of this Creative Commons license, please visit [creativecommons.org](http://creativecommons.org).

If you have any questions about citing or reusing New America content, please visit [www.newamerica.org](http://www.newamerica.org).

All photos in this report are supplied by, and licensed to, [shutterstock.com](http://shutterstock.com) unless otherwise stated. Photos from federal government sources are used under section 105 of the Copyright Act.