

ENVISIONING THE NATIONAL POSTSECONDARY DATA INFRASTRUCTURE IN THE 21ST CENTURY

# Creating a Thriving Postsecondary Education Data Ecosystem

**ARCHIE CUBARRUBIA**

MIAMI DADE COLLEGE

**PATRICK PERRY**

WESTED

MAY 2016

**Archie Cubarrubia**, Ed.D., is vice provost for institutional effectiveness at Miami Dade College. **Patrick Perry** is chief information officer for the California State University Chancellor's Office. From May 2015–April 2016, he served as senior research associate at WestEd.

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, the Institute for Higher Education Policy, Miami Dade College, the California State University, or WestEd.

# Creating a Thriving Postsecondary Education Data Ecosystem

Leveraging data reported regularly by colleges and universities into useful information is critical to ensuring that more students have access to and succeed in higher education. However, the needs of the higher education community—from state and federal policymakers, to faculty and staff at institutions, to students themselves—are evolving and growing in complexity. The current national postsecondary data infrastructure, with its component systems built for different purposes and owned and operated by different entities using different data definitions, is increasingly inadequate to meet those needs. A variety of entities in the U.S. today produce and report a staggering amount of data on postsecondary students, their enrollment behaviors and patterns, and their eventual outcomes. Yet, significant gaps in the data on which we as postsecondary education researchers and stakeholders currently depend leave important questions largely unanswered:

- ▶ How many part-time, adult, military, remedial, and Pell Grant students are successfully completing their degrees?
- ▶ Are students who do not graduate transferring to another school or dropping out altogether?
- ▶ How much are students learning in college, and how are they contributing to society after college?
- ▶ Are students leaving school with loan debt, and are they paying back their loans, especially students who do not complete?
- ▶ What types of jobs are students getting after college? Are they going to graduate school?<sup>1</sup>

Currently, postsecondary data are collected and shared at many levels and locations: at the postsecondary institution, within states, in multistate collaboratives, in privately held databases, and at the federal level. Each of these repositories represents a piece of our national postsecondary data infrastructure. Evolving this infrastructure—a fragmented “system” that somewhat haphazardly collects, aggregates, and reports educational data—into a coordinated ecosystem can greatly enhance our ability to leverage postsecondary data to answer these and other questions to improve student success.

Creating a coordinated ecosystem requires intentional change. Removing existing legal barriers and reducing the burden of data collection can simplify the task of accessing and using data at all levels of collection and reporting. Imple-

menting standardized, common data definitions and formats can not only improve the usefulness of reporting, but also provide datasets able to respond to a multitude of questions. Aggregating postsecondary data to higher levels, including nationally, can provide a backbone structure able to improve student success and close equity gaps. These and other critical improvements comprise the core of recommendations from this paper series.

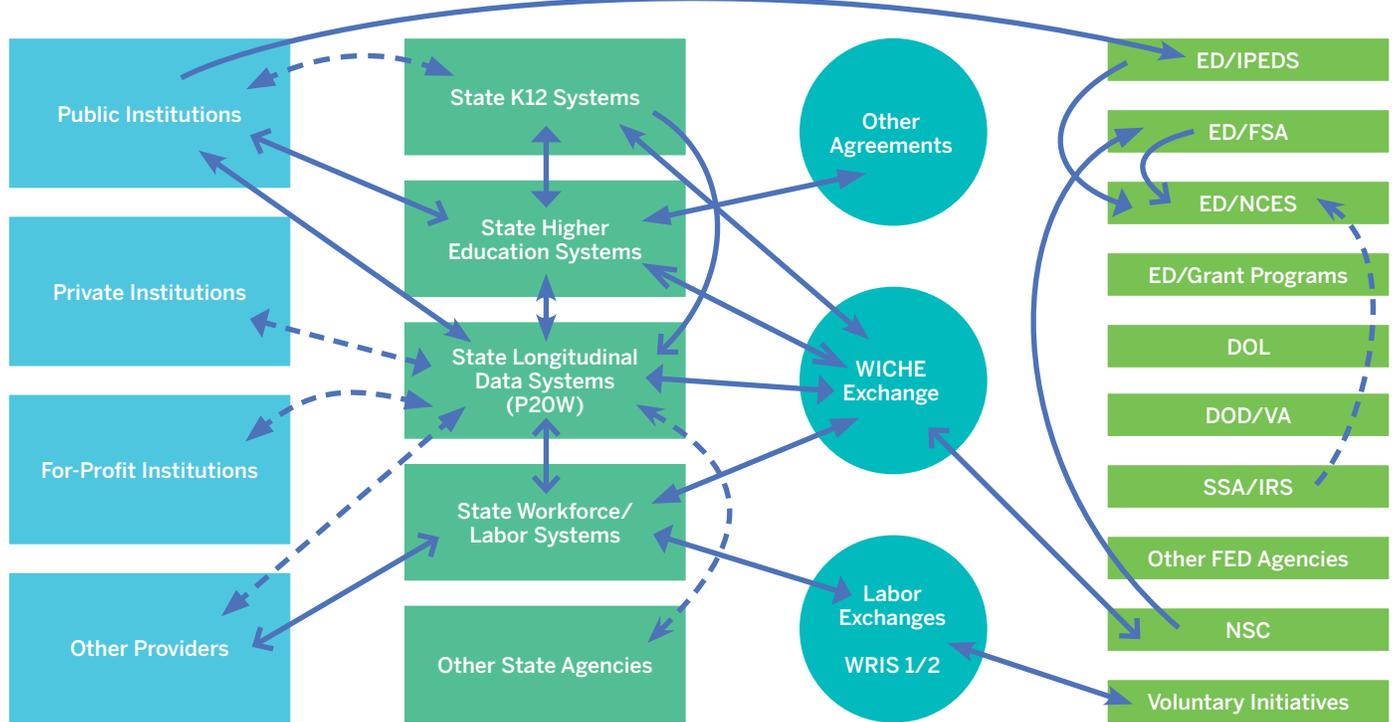
Leveraging the Postsecondary Data Collaborative (PostsecData), the Institute for Higher Education Policy (IHEP) convened an expert working group to develop technical and policy options for improving the national postsecondary data infrastructure to inform state and federal data policy conversations around student success. The working group members are composed of experienced and influential practitioners and policymakers who have worked with the existing infrastructure for years. Their recommendations represent proposals to improve existing components of the national postsecondary data infrastructure, or in some cases, create new ones. Many of these recommendations can be easily put into practice today, while others portend a more complex path to implementation. But none point specifically to any one single solution as a silver bullet. The goal is to move all current systems toward a more cohesive and interoperable postsecondary data ecosystem.

This paper frames the work of these national experts and makes the case for engaging in this conversation *now*. Improvements to which data we collect from colleges and universities, how they are collected, where they are aggregated and stored, how they can be linked to other existing data, and how they are disseminated to the public are all within reach.

## Assessing the Current State of the National Postsecondary Data Infrastructure

Currently, there is not a singular, well-planned, and coordinated national postsecondary data system. Instead, there is a patchwork of individual data systems built for different purposes, governed by different statutes and regulations, owned and operated by different entities, and using different data definitions (see **Figure 1**). At best, users can access and use certain data from each system to build a somewhat coherent, if incomplete, national picture of postsecondary education. We can measure the success of certain students, the success of certain institutions, and the success of certain taxpayer

**FIGURE 1: CURRENT POSTSECONDARY DATA INFRASTRUCTURE—INCOMPLETE, DUPLICATIVE, DISCONNECTED SYSTEMS INCREASE BURDEN, DECREASE UTILITY**



investments in the enterprise. At worst, the existing infrastructure fails to count a significant number of students and institutions and facilitates redundancy, collection burden, waste, and poor data quality that hinder us from achieving our most important mission: promoting student success and educational equity in college degree attainment.

It should be noted, however, that although the current infrastructure may seem fractured, imperfect, and incomplete, it is not because it involves many players. Simply calling for a single, unified system in and of itself is not a feasible solution. It is unrealistic to envision a future in which multiple data repositories—each built for and fulfilling unique purposes—do not coexist. What will represent an improvement over today’s postsecondary data infrastructure is optimization of these data repositories that allows actors and systems to function more efficiently with each other and within a larger ecosystem by reducing redundancies and burden and increasing access to more useful data.

The data that form the backbone of these various, disconnected systems start with students as they begin their application process into the postsecondary system. As they move into and through higher education institutions, data on their enrollment, financial aid, progress, completion, and outcomes are appended. These data are then aggregated at

higher levels and reported by institutions, federal or state governments, and private organizations to populate reports such as segmental and state accountability dashboards, the U.S. Department of Education (ED) College Scorecard, and the Federal Student Aid (FSA) Data Center.

Generally speaking, data in the current postsecondary data infrastructure flow as follows:

- ▶ Students supply information about themselves, their demographics, prior academic histories, and academic goals on applications to individual colleges and universities and on financial aid applications to federal and state governments.
- ▶ As students progress, data on their enrollments and outcomes are captured by institutions, along with their financial history and needs and information on the postsecondary credentials they earn.
- ▶ Many institutions submit these individual student records to state or other intermediary coordinating boards and inter/intrastate educational data repositories and clearinghouses, forming multi-institution data warehouses. Some warehouses also link this postsecondary information to data on students’ K-12 pathways and workforce outcomes within or across states.

- ▶ Almost all institutions voluntarily submit a subset of unitary student data to the National Student Clearinghouse (NSC). NSC data are collected for a variety of purposes that assist institutions, primarily with the tracking and verification of students as they move to other institutions.<sup>2</sup>
- ▶ As a condition of participation in federal student aid programs, institutions also submit student-level data to FSA and aggregated institution-level data to the National Center for Education Statistics (NCES) on a core set of required metrics as part of the Integrated Postsecondary Education Data System (IPEDS). However, the federal government does not act as a single, national repository of student-level data for all students, but only for students who receive federal Title IV aid.
- ▶ States and federal governments also collect information on employment and wages, typically from employers or other government agencies. These data can be matched to students based on a common identifier like a Social Security number. Access to personally identifiable information is extremely limited and governed by strict privacy laws, most notably the Family Educational Rights and Privacy Act (FERPA); Joanna Lyn Grama’s paper in this series provides more detail on this subject).<sup>3</sup>

Stakeholders use data in the current infrastructure for valid and varied reasons. For example:

- ▶ The federal government uses data to ensure that over \$150 billion in student aid programs are accessed and repaid, to promote institutional accountability across institutions whose students receive financial aid, and to provide consumer information on the types of institutions and programs that exist, their costs, and the success of students enrolled.
- ▶ State governments, educational coordinating boards, and segmental offices use data to perform institutional and program accountability, to allocate and apportion funding to institutions, to administer financial aid programs, to advise in policymaking, to inform students and the public, and to support data-driven systems of continuous improvement.
- ▶ Colleges and universities use these data to compare performance against peer and aspirational institutions and to assess progress on student success and institutional effectiveness efforts.
- ▶ Researchers, policy advocates, and philanthropic organizations use data collected by state and federal governments, by institutions, and by private organizations to support specific initiatives and to perform program evaluations of targeted interventions.

While this may seem to be a well-coordinated channel of data flowing to these various entities, the functional limitations of the existing national postsecondary data infrastructure are well known to researchers, policymakers, and institutions themselves. It is instead a fractured system, beset with legal governance and stewardship limitations, non-standardized data definitions, formats and structures, and few accessible aggregated data stores that can answer a wide array of very important questions about the higher educational enterprise today. While each stakeholder—federal, state, institutional, or private—does their best to answer the questions that can be answered with the datasets available to them, each has limitations on what questions they can answer. More critical, none have a clear path to accessing the comprehensive set of data that the other component systems contain.

These papers, each covering a specific component of the national postsecondary data infrastructure, identify areas for improvement and provide recommendations. Their content is not meant to compete with each other; they are rather all offering complementary approaches that aim to improve data quality, transparency, and use while ensuring security and privacy. To reiterate, there is no silver bullet. Ultimately, what will make a difference is a coordinated and strategic approach to how each of these existing data systems—and their respective stakeholders—can work together to answer the critical questions being asked now about higher education, not to mention the critical questions that will be asked of us twenty years from now. Each stakeholder plays a role. But there is much potential to reduce costs and burden associated with redundancy and poor data quality, and produce better information in service of better student outcomes. It is beyond the time to modernize our disparate postsecondary data systems to create an agile and effective information ecosystem for higher education.

## Creating an Agile and Effective Postsecondary Data Ecosystem

### Guiding Principles

We recommend the following principles to guide the work of developing an agile and effective postsecondary data ecosystem:

- ▶ Any improvement to the national postsecondary data infrastructure must advance student equity and success. As the expert working group engaged in conversations regarding improvements to various data systems, we asked ourselves a key question: To what end? We cannot underscore enough the importance of improving our national postsecondary data infrastructure to produce information that will be used to substantially improve access and success, particularly for historically underrepresented students. Data system improvements are only worthwhile if they can translate into better use of data to improve student outcomes.
- ▶ Improvements must result in increased efficiency in data collection, reporting, and use. Institutional effort in meeting required and voluntary reporting is not inconsequential. However, when evaluating burden, this effort should be weighed against the value of the data to policymakers, to students, and to institutions themselves. Therefore, the ultimate goal of this process should not be to decrease institutional effort at large (although reducing duplication of effort will likely decrease burden), but to determine acceptable levels of effort appropriate to increasing the value of the information reported.
- ▶ Improvements must result in flexibility and agility that allows the ecosystem to address data needs today and twenty years from now. Continuing to address questions about the higher education enterprise with siloed data collections designed to inform singular issues is a practice that has led to today's fractured data environment, along with ever-increasing institutional burden. While we cannot yet determine what questions will need to be addressed in the future—just as twenty years ago we could not have predicted the advent of online delivery modes and competency-based education—what we can do now is create flexible, connected data systems so that new insights and knowledge can be easily culled without further taxing limited institutional, state, and federal resources.
- ▶ Creation of an agile and effective postsecondary data ecosystem cannot come at the expense of the privacy of the individuals whose personal and educational records are contained within it. Best and leading practice in education and other industries should be applied to ensure that individuals' data remain private and secure to meet and exceed statutory and regulatory requirements.

### Papers in this Series

The papers in this series propose improvements to components of the national postsecondary data infrastructure and provide a roadmap for intentionally evolving the current infrastructure into a thriving ecosystem. The papers generally organize proposed improvements into the following categories:

- ▶ **Enhancing existing systems.** Improvements to existing data collection and reporting systems currently in use to collect, store, and disseminate educational data can and should be pursued now, under each system's existing authority. Jamey Rorison and Mamie Voight propose improvements to the Integrated Postsecondary Education Data System (IPEDS). Matthew Soldner and Colleen Campbell propose improvements to FSA data systems. Afet Dundar and Doug Shapiro propose improvements to the National Student Clearinghouse. John Armstrong and Katie Zaback propose improvements to state postsecondary student unit record systems.
- ▶ **Improving interoperability.** Interoperable systems provide for better information and reduce duplication of effort. Building this interoperability, such as by creating the ability for datasets housed in federal agencies to be connected and leveraged, makes sense. David Bergeron proposes streamlining and coordinating data collections across federal agencies outside ED. Interoperability across federal, state, and privately held systems may be a bigger challenge but should also be examined. To strengthen these linkages, Brian Prescott and Patrick Lane propose improvements to state-to-state data exchanges, and Rachel Zinn proposes improvements to linking data on workforce outcomes in state and federal systems.
- ▶ **Developing new systems.** In some cases, gaps in information on student access, costs, and outcomes can be addressed by developing new systems that link together and supplement existing data. A single, unitary data collection across all states and segments of higher education can provide a useful and consistent source of data to support policymaking, decision support, accountability, and consumer information. Ben Miller explores options for accomplishing these goals through the development of a federal student-level data system.
- ▶ **Improving data quality and use.** Because data are ultimately reported by individuals at colleges and universities, quality is contingent on the professional capacity and expertise of those individuals. Randy Swing proposes improvements to the field of institutional research that can also be applied to other professions involved in data reporting. Finally, Joanna Lyn Grama provides best practices for ensuring data privacy and security, a critical consideration throughout the postsecondary data ecosystem.

Across the papers, authors included a core set of recurring recommendations for creating an agile and effective national postsecondary data ecosystem:

- ▶ **Improve the capacity of institutions to report accurate, timely, and relevant data.** An underlying assumption of the papers' recommendations is that the data that serve as the lifeblood to each of the existing data systems are accurate, timely, and relevant. If institutions—as the primary producers of these data—are not equipped to provide quality data, structural improvements to the ecosystem will be for naught.
- ▶ **Collect new data elements; eliminate others.** Many papers call for adding elements to or deleting them from existing data collections. For example, several papers call for creating a single national postsecondary institutional identifier, adding information on job classifications, and identifying types and hours of work on unemployment insurance wage records. Also, several recommendations propose to eliminate data elements that are no longer widely used. Based on a thorough review of the data elements used in 20 different data collection initiatives, IHEP produced a technical metrics guide that proposes common, consistent definitions across the entire ecosystem.<sup>4</sup> Use of common definitions could reduce duplication and institutional reporting burden and allow for the development of new metrics and performance indicators that would be useful to policymakers, consumers, and institutions.
- ▶ **Improve access to data.** One of the main barriers to creating processes of continuous improvement is the inability to access much of the data in the ecosystem by both researchers and by individual institutions. Improvements can be made by developing better public-facing business intelligence tools, by allowing institutions access to additional information (e.g., national wage data, enrollments at other institutions), and by creating a “researcher access” restricted user license to certain data sets. In all cases, higher education stakeholders strongly desire an improvement in the ability to access disaggregated data across multiple variables.
- ▶ **Keep individual data private and secure.** The papers stressed the importance of creating systems, processes, and policies that keep data private and secure. This is important across all levels and aggregations, with explicit emphasis on individual data. All data use must comply with the appropriate laws and should follow industry standard best practices.

- ▶ **Improve linkages between existing datasets.** Many existing datasets have the capability of linking together to answer a wide range of questions. These include federal IPEDS, FSA, Department of Defense, Veterans Affairs, Social Security Administration, National Directory of New Hires, and Census databases; state postsecondary student unit record systems and state Employment Development Department wage databases, including the Wage Record Interchange System (WRIS) and the Wage Record Interchange System 2 (WRIS2); NSC data; and segmental and institutional databases. Hindering these linkages are a variety of institutional risk-aversions, legal barriers (perceived or otherwise), misaligned matching key fields, and fiscal considerations. Much of the data necessary to create an agile and effective data ecosystem exist; breaking down the silos and enabling systems to connect to each other is a high priority in many of the papers.
- ▶ **Remove legal barriers to data use.** Authors in this series consider eliminating real or perceived legal obstacles to collecting, sharing, storing, and reporting data as worthy of effort. This includes revising definitions of who is considered an “authorized representative” allowed to access data; specifically authorizing data matching between and among federal and state agencies; allowing for institutional access to specific datasets; and calling for the repeal of the ban on the creation of a federal postsecondary student unit record data system. It is important to note that the paper authors are not suggesting a relaxation of data privacy and security regulations or protocols; the goal is to allow for a more efficient flow of data while protecting student privacy.

A summary of the papers and their respective recommendations is located in **Table 1**. These recommendations span shorter-term, more easily implemented solutions that can result in immediate impacts and longer-term, more challenging, and sometimes aspirational options that will require time, money, and effort. However, we caution against enacting only “quick fixes”; comprehensive solutions are what will ensure that we do not have these same conversations regarding postsecondary data twenty years from now.

**TABLE 1. PAPER SERIES SUMMARY AND RECOMMENDATIONS**

Title and Authors	Summary	Primary Recommendations
<p><b>Assessing and Improving State Postsecondary Data Systems</b></p> <p>John Armstrong, State Higher Education Executive Officers Association Katie Zaback, Complete College America</p>	<p>Most states have postsecondary student unit record systems (PSURS) that combine student records from across a student’s entire educational career into the workforce. Improving these will help strengthen states’ abilities to develop informed policies.</p>	<ul style="list-style-type: none"> <li>▶ Invest in state PSURS, and adopt best practices from other state-level unit record systems.</li> <li>▶ Involve the state PSURS in the postsecondary agency’s strategic plan for higher education.</li> <li>▶ Continue to expand the use of state PSURS and communicate their value.</li> <li>▶ Proactively address privacy and security concerns; perform additional outreach to policymakers to ensure data security protects against unnecessary restrictions that prevent meaningful research.</li> <li>▶ Continue to fund state data systems federally while shifting the focus of federal grants to outcomes rather than infrastructure.</li> <li>▶ Allow state PSURS access to federal datasets to improve matching and linkages.</li> <li>▶ Enable better state-level analysis via federal datasets.</li> <li>▶ Use lessons from recent state PSURS improvements if and when developing a federal unit record system.</li> </ul>
<p><b>Building a Student-Level Data System</b></p> <p>Ben Miller, Center for American Progress</p>	<p>Currently banned by federal legislation, the creation of a federally run national postsecondary student unit record data collection has tremendous promise for answering today’s and tomorrow’s policy questions. The paper explores multiple options for creating such a system.</p>	<ul style="list-style-type: none"> <li>▶ Remove the ban on developing a federal student unit record system.</li> <li>▶ Convene technical review panels for necessary new indicators.</li> <li>▶ Improve alignment between institutional research and financial aid reporting.</li> <li>▶ Add data dissemination and disclosure to FSA’s organizational goals.</li> <li>▶ Modernize FSA’s main student database.</li> <li>▶ Create a researcher license for FSA data.</li> </ul>
<p><b>Classroom to Career: Leveraging Employment Data to Measure Labor Market Outcomes</b></p> <p>Rachel Zinn, Workforce Data Quality Campaign</p>	<p>The wage and workforce outcomes of students are helpful for promoting institutional program development, accountability, and student choice. Improving access to these data for a variety of stakeholders is paramount.</p>	<ul style="list-style-type: none"> <li>▶ Allow federal agencies to match student records with employment data and display employment and earnings outcomes by program of study.</li> <li>▶ Amend the Family Educational Rights and Privacy Act (FERPA) to include provisions to allow an “authorized representative” to evaluate education programs, broadly defined.</li> <li>▶ Use Workforce Investment and Opportunity Act (WIOA) regulations to clarify permissible purposes and parties to access UI wage records, including state education agencies and colleges.</li> <li>▶ Issue joint ED/Department of Labor (DOL) guidance to promote data linkages and uses, and clearly explain how state data systems may link education and wage data in compliance with FERPA and UI rules.</li> <li>▶ Push Congress to support federal funding for states to enhance their data systems, including linkages of education and wage data.</li> <li>▶ Enact state policies that promote transparency on employment outcomes of postsecondary programs and the data systems required to calculate them.</li> <li>▶ Amend state laws and restrictive legal opinions that unnecessarily inhibit wage data access.</li> <li>▶ Improve Wage Record Interchange Systems (WRIS, WRIS2) and Federal Employment Data Exchange System (FEDES) data sharing agreements to include all states and expand the terms of data usage to include the evaluation of employment outcomes.</li> <li>▶ Create interagency data governance councils, data sharing agreements, and staff training protocols to build trust, ensure confidentiality and security, and develop a culture of data sharing and use.</li> <li>▶ Allocate state funds to maintain and improve employment data linkages and support use of linked state data.</li> <li>▶ Pilot efforts to enhance UI wage records, including hours worked and occupational codes, to make the data more valuable for assessing labor market success.</li> </ul>

Title and Authors	Summary	Primary Recommendations
<p><b>Fostering State-to-State Data Exchanges</b></p> <p>Brian T. Prescott, National Center for Higher Education Management Systems</p> <p>Patrick Lane, Western Interstate Commission for Higher Education</p>	<p>Linking existing state longitudinal data systems in a “federated” model, where data can be accessed across states while managed within each state, is currently being piloted. This approach might provide a feasible alternative for creating more access to postsecondary data in a decentralized model.</p>	<ul style="list-style-type: none"> <li>▶ Elevate cross-state data partnerships as a priority.</li> <li>▶ Fund analytical work that shows the effects of student mobility on outcomes.</li> <li>▶ Allow states to access federal data systems to analyze employment outcomes and evaluate programs.</li> <li>▶ Require submission of additional data elements to unemployment insurance (UI) wage records.</li> <li>▶ Promote metrics that capture student outcomes at other institutions and in employment in other states.</li> <li>▶ Ensure that state data systems include elements that are used in evaluating equity and employment outcomes.</li> <li>▶ Require employers to submit additional data elements as part of their UI data submissions.</li> <li>▶ Ensure that data elements necessary to connect education and employment are available for use.</li> <li>▶ Allow institutions to use employment data from state UI wage records at the unit level.</li> </ul>
<p><b>Institutional Research Capacity: Foundations of Federal Data Quality</b></p> <p>Randy L. Swing, Association for Institutional Research</p>	<p>Cutting across the issues related to state, federal, and national data collection and reporting is the human capacity located at campuses to submit and use data to improve student success and institutional effectiveness.</p>	<ul style="list-style-type: none"> <li>▶ Establish an intentional data strategy for the overall postsecondary data ecosystem and for each component of the ecosystem.</li> <li>▶ Build in disaggregation capacities while planning data collections so that data can be useful in decisions at tactical, operational, and strategic levels.</li> <li>▶ Establish at each institution a Chief Institutional Research Officer (CIRO) as a cabinet-level position.</li> <li>▶ Develop an intentional institutional plan for staff professional development of data literacy skills.</li> <li>▶ Require federal statistical agency missions include authority to train data providers and data consumers in their respective roles in the data ecosystem.</li> <li>▶ Ensure federal calculations of reporting burden use a cost–benefit approach that acknowledges the value of data used by the reporting sources and to the federal government.</li> <li>▶ Give equal consideration when designing data collections to the automation, distribution, and use of the data.</li> <li>▶ Allow institutional-level data strategies to take advantage of disruptive innovations already in play and update their strategies as new technologies become available.</li> </ul>

Title and Authors	Summary	Primary Recommendations
<p><b>Leveraging What We Already Know: Linking Federal Data Systems</b></p> <p>David A. Bergeron, Center for American Progress</p>	<p>There are any number of separate federal data collections such as the National Directory of New Hires (NDNH), the Social Security Administration's (SSA) wage and earnings data, the Internal Revenue Service's tuition and required fees and financial aid data, the DoD's military recruiting data and data from the Department of Veterans Affairs (VA) that have great potential of adding to the domain of knowledge of higher education. Successfully enabling these linkages legally and politically is critical to making this happen.</p>	<ul style="list-style-type: none"> <li>▶ Demonstrate what we do not know about the outcomes of our educational investments at the point where students and families are making decisions about postsecondary education institutions and programs.</li> <li>▶ Demand greater transparency from institutions for information on outcomes for graduates.</li> <li>▶ ED must play a leadership role in demonstrating the benefits of sharing data across governmental agencies.</li> <li>▶ Repeal the ban on a student unit record system and allow for linkages to additional federal data resources in a federal student unit record system.</li> <li>▶ Provide access to earnings data to the Secretaries of ED and VA to assess the economic impact of institutions and programs.</li> <li>▶ Consolidate the NDNH with SSA's wage and earnings reporting.</li> <li>▶ Provide a statutory exception to the Privacy Act of 1974 to provide for the exchange of federal data on individuals for the purposes public disclosure and evaluation of student financial aid programs.</li> <li>▶ Explore the benefits to NDNH in matching their data with education data to aid in the enforcement of child support orders.</li> <li>▶ Explore whether state longitudinal data systems could be enhanced by NDNH data.</li> <li>▶ Consider entering into an agreement with SSA to obtain aggregate earnings data used to calculate return on investment (ROI).</li> </ul>
<p><b>The National Student Clearinghouse as an Integral Part of the National Postsecondary Data Infrastructure</b></p> <p>Afet Dunder and Doug Shapiro, National Student Clearinghouse Research Center</p>	<p>There currently exists a national, nongovernmental longitudinal student unit record data collection with voluntary participation by postsecondary institutions. The National Student Clearinghouse (NSC) has any number of key advantages in collecting and using data, but some limitations due to its voluntary participation policy and the data are not available to the public or policymakers.</p>	<ul style="list-style-type: none"> <li>▶ Encourage and expand existing incentives for voluntary institutional transparency; develop metrics that allow for benchmarking institutional improvement.</li> <li>▶ Build stronger public-private partnerships among institutions, ED, and NSC to provide enhancements to public data such as IPEDS with minimal increase in institutional burden.</li> <li>▶ Develop a public access data facility that would bring PowerStats-like functionality to NSC data.</li> <li>▶ Extend existing linkages between NSC data and local and state data systems.</li> </ul>
<p><b>Putting the "Integrated" Back Into IPEDS: Improving the Integrated Postsecondary Education Data System to Meet Contemporary Data Needs</b></p> <p>Jamey Rorison and Mamie Voight, Institute for Higher Education Policy</p>	<p>As the primary federal data collection on postsecondary institutions, the Integrated Postsecondary Education Data System (IPEDS) compiles useful information but is limited in scope because it is an aggregate, rather than student-level, collection. Improving the current data collection can act as an intermediate step until a more robust and flexible solution is implemented.</p>	<ul style="list-style-type: none"> <li>▶ Collect data on key performance indicators that fully reflect today's students: improved graduation rate, retention rate, levels of academic preparation, gateway course success, units upon completion.</li> <li>▶ Streamline IPEDS to reduce burden and produce more relevant data; report usage rates of survey data to evaluate underused data elements.</li> <li>▶ Create linkages between IPEDS and other systems, including Federal Student Aid (FSA), the Department of Defense (DoD), the Department of Veterans Affairs (VA), the Department of the Treasury, and the U.S. Census Bureau.</li> <li>▶ Create a single, national, campus-level identifier for every institution of higher education.</li> <li>▶ Use IPEDS data to simplify institutional reporting on federal grants and to prepopulate Title III and Title V Annual Performance Reports.</li> </ul>

Title and Authors	Summary	Primary Recommendations
<p><b>Understanding Information Security and Privacy in Postsecondary Education Data Systems</b></p> <p>Joanna Lyn Grama, EDUCAUSE</p>	<p>The laws and ethics surrounding data security and privacy cut across all levels of data collection, sharing, access, and reporting.</p>	<ul style="list-style-type: none"> <li>▶ Adopt a risk-based approach to understanding information security and privacy threats and vulnerabilities.</li> <li>▶ Establish and adhere to a baseline set of information security protections.</li> <li>▶ Establish and adhere to a baseline set of privacy standards.</li> <li>▶ Implement a collaborative governance structure that includes addressing information security and privacy throughout the national postsecondary education data infrastructure.</li> </ul>
<p><b>Using—and Improving—FSA Data Systems to Support Policy Analysis</b></p> <p>Matthew Soldner, American Institutes for Research</p> <p>Colleen Campbell, Association of Community College Trustees</p>	<p>Federal Student Aid (FSA), an office of the U.S. Department of Education (ED), collects data on students who apply for and receive federal student aid. The transactional data collected through FSA's systems can be used more for analytical purposes.</p>	<ul style="list-style-type: none"> <li>▶ Develop a feedback process on the types of FSA data-based analyses that would benefit policymakers and analysts, and post the resulting information to the FSA Data Center.</li> <li>▶ Improve the usefulness of existing National Student Loan Data System (NSLDS)-based reports provided to campus-based aid administrators and ensure it supports local research designed to improve management of aid programs.</li> <li>▶ Build the capacity of the nascent FSA Data Office to respond to more complex research requests.</li> <li>▶ Leverage ED's existing PowerStats web tool to allow for the analysis of extracts from FSA data systems.</li> <li>▶ Use ED's existing restricted-use data licensing process to make FSA data extracts or ED policy analysis tools (e.g., Pell Estimation Model) available to qualified researchers.</li> <li>▶ Explore whether access to ED's new Enterprise Data Warehouse might be possible under the Census Bureau's Research Data Center model.</li> </ul>

## Cross-Cutting Considerations

A set of recurring considerations undergird the papers in this series. These include the following:

### Data Governance

A key consideration in building an agile and effective national postsecondary data ecosystem is governance: who owns these data and the systems that collect, analyze, and disseminate them? Because current systems have been built to meet specific purposes, improvements do not necessarily require transferring ownership of the system—although in some cases, it may be more efficient to do so. The success of improvement efforts is largely dependent on the willingness and ability of system owners to work together. In addition, data collectors, stewards, and their governance structures will need to address technical considerations related to system interoperability.

### Data Standards and Definitions

Adding complexity to the discussion is the use of different data standards and definitions by different data collection entities. For example, nuances around which students are included in the denominator of a particular graduation rate metric for federal purposes versus state purposes may be important to the respective policymakers but may be lost to students and families deciding on which institution is a better fit. Stakeholders in the field must address the challenges related to developing a common, consistent set of data standards and definitions across the different components of an agile and effective data ecosystem. The Common Education Data Standards (CEDS) initiative and the field-driven Postsecondary Metrics Framework<sup>5</sup> are steps in the right direction.

### Data Use

Defining the allowable uses of the data is important to structuring data collection and sharing and is critical in building relationships between existing systems. In addition, improving the ability for data to be reported to the public and the community of practice is an important consideration. Promoting institutions' use of data for improvement and accountability purposes is also critical, since the value of the data ultimately relies on campus leaders to advance a culture of continuous, data-driven improvement.

### Data Policy

State and federal statutes and regulations can be barriers to innovation—or they can promote the collection and use of high-quality data. Some improvements to existing data systems can be accomplished within agencies' existing authorities. Others—such as the creation of a federal student unit record system—require statutory change. In many cases the proposed improvements require political will and vision.

## Data Resources

As expressed earlier, burden at all levels must be acknowledged and managed based on an intentional analysis of costs and benefits. The papers consider explicit costs of data sharing/aggregation/warehousing, as well as potential economies of scale. The various recommendations in this series will have various associated costs, each of which should be weighed against the benefits. But many improvements may have upfront costs that will decrease over time, even for more comprehensive system overhauls, as reporting becomes routine practice.

### Privacy, Security, and Institutional Research Capacity

Privacy and security as well as institutional capacity to produce quality data are two cross-cutting considerations that have been given their own separate platforms in this series because both are imperative to the success of proposed improvements to each individual system and to the larger ecosystem. Both considerations span and weigh against the collective recommendations.

*Privacy and security.* Safeguarding individuals' data by keeping them private (with no individual able to be identified personally) and secure (unable to be accessed by non-authorized parties) is nonnegotiable. Critical to these conversations is determining which individuals at which entities should have access to which data and for which purposes. Best practices to protect privacy and security are already in use in education and in other industries such as healthcare, and a legal framework of current statutes and regulations that allow for specific yet safe uses of student data does not exist.

Different approaches have different privacy and security implications. For example, an aggregate, institution-level data collection such as IPEDS is different from systems that collect student-level data such as state postsecondary unit record systems or the National Student Clearinghouse. State postsecondary student unit record systems are governed by their respective state statutes and regulations, which have varying levels of restrictions around who can access the data and for what purpose, in addition to federal mandates. Therefore, data sharing across states to ascertain outcomes of students enrolled at border institutions, for example, requires careful negotiation between state agencies for data access and use that complies with the law. Similarly, wage data from state unemployment insurance databases are governed by different statutes and regulations than are wage data from the Social Security Administration and as such have varying implications for access and use.

*Institutional capacity to produce quality data.* The most agile and effective data ecosystem is only as good as the data that populate it. Although quality control is a feature of each system, responsibility for collecting, preparing, and reporting data still falls on individuals on each campus, and data quality depends on those individuals' understanding of each system's reporting requirements and the definitions of the data elements within each. Nearly all accredited colleges and universities that participate in Title IV programs have established an Institutional Research (IR) function that supports state, federal, and other external reporting on enrollments, resources, and student outcomes. Yet the variation in investments to the IR function creates vast differences in the capacity of institutions to produce accurate, timely, and relevant data to comply with mandated reporting and to support data-driven decision making. It is also worth noting that state and federal reporting is not the exclusive domain of institutional research. Rather, financial aid officers, registrars, and others hold data reporting responsibilities as well. Therefore, as higher education data stakeholders undertake efforts to improve state and national data systems, they must also improve institutions' data reporting capacity—whether in institutional research offices or elsewhere on campus—in order to ensure data quality. Otherwise, improvements to the ecosystem will be in vain.

## Conclusion

The goal of this paper series is to inform policymakers at the local, state, and federal levels about how we can build an agile and effective postsecondary education data ecosystem that will meet current, emerging, and future needs. The recommendations in these papers are part of a broader effort to improve our ability to use data to facilitate student success and close equity gaps and therefore should be considered complementary to the promising work already underway.

The number and complexity of questions about the effectiveness of our collective investment in higher education continue to grow. Policymakers, advocates, and even institutions themselves are unable to answer these questions using the existing national postsecondary data infrastructure, which is burdensome, uncoordinated, and increasingly at risk of slipping into obsolescence. To move forward as a nation, we must take the opportunity now to create an agile and effective national postsecondary data ecosystem whose individual components communicate with and build upon each other to enable all stakeholders in the enterprise to focus on what really matters: student success.

## Endnotes

---

- 1 Voight, M., Long, A., Huelsman, M., and Engle, J. (2014). *Mapping the postsecondary data domain: Problems and possibilities*. Washington, DC: Institute for Higher Education Policy. Retrieved from [http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/mapping\\_the\\_postsecondary\\_data\\_domain.pdf](http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/mapping_the_postsecondary_data_domain.pdf)
- 2 National Student Clearinghouse. (2016). Our history. Retrieved from [http://www.nslc.org/about/our\\_history.php](http://www.nslc.org/about/our_history.php)
- 3 U.S. Department of Education. (2016). Family Educational Rights and Privacy Act (FERPA). Retrieved from: <http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>
- 4 Janice, A. and Voight, M. (2016). *Toward convergence: A technical guide for the Postsecondary Metrics Framework*. Washington, DC: Institute for Higher Education Policy. Retrieved from <http://www.ihep.org/postsecdata/mapping-data-landscape/metrics-framework-technical-guide>
- 5 Engle, J. (2016). *Answering the call: Institutions and states lead the Way toward better measures of postsecondary performance*. Seattle, WA: The Bill and Melinda Gates Foundation. Retrieved from <http://postsecondary.gatesfoundation.org/wp-content/uploads/2016/02/AnsweringtheCall.pdf>







*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.

