A Blueprint for Better Information:

Recommendations for a Federal Postsecondary Student-Level Data Network

AMANDA JANICE ROBERSON, JAMEY RORISON, AND MAMIE VOIGHT

October 2017

A REPORT BY
Institute for Higher Education Policy

SUPPORTED BY
The Bill & Melinda Gates Foundation

IHEP

Acknowledgments

This report is the result of hard work and thoughtful contributions from many individuals and organizations. We would like to thank the Institute for Higher Education Policy staff who helped in this effort, including Michelle Asha Cooper, president; Kelly Leon, communications and advocacy officer; Lacey Leegwater, senior advisor; Alain Poutré, research analyst; Katherine Wheatle, research analyst; and India Heckstall, Tyler Wu, and Jack Porter, former policy interns.

We are appreciative of Sanametrix, specifically Joe Collins, Zac Mangold, Christina Tellez, Jerry Malitz, and Brett Richards, who created the technical framework for this project. We also thank a number of experts and organizations who participated in the advisory committee, including:

Matthew Soldner, American Institutes for Research, Committee Chair
Kazim Ali, Formative Co.
Sherry Bennet, University of Maryland University College
Courtney Brown, Lumina Foundation
Cole Clark, Deloitte
Afet Dundar, National Student Clearinghouse
Jennifer Engle, Bill & Melinda Gates Foundation
Tamar Epstein, Ellucian
Neal Gibson, formerly of the Arkansas Research Center
Will Goldschmidt, DBDriven.Net LLC
Joanna Lyn Grama, EDUCAUSE
Terrell Halaska, HCM Strategists
Ron S. Jarmin, U.S. Census Bureau
Christine Keller, Association for Institutional Research
Amy Laitinen, New America
Patrick Lane, Western Interstate Commission for Higher Education
Nicole Melander, Civitas Learning
Ben Miller, Center for American Progress
Doug Newhard, Hartman Executive Advisors
Monal Patel, Purdue University
Neil Ridley, Georgetown University Center on Education and the Workforce
Jeff Sellers, AEM Corporation
Douglas Shapiro, National Student Clearinghouse
Kathy Stack, Laura and John Arnold Foundation
Rachel Zinn, Workforce Data Quality Campaign

Finally, we are grateful for the Bill & Melinda Gates Foundation's support of this and other research on postsecondary data. Although many have contributed their thoughts and feedback throughout the production of this report, the research and recommendations presented here are those of the authors alone.

Executive Summary

Research shows that investing in a college education pays dividends for both students and society.[1] But, persistent gaps in college access, success, and attainment prove that not everyone shares the same benefits. Outcomes vary within and across institutions, yet we often don't know which programs at which institutions provide a return on investment, and for which students. The data that students, institutions, and policymakers currently use for decision-making are disconnected, duplicative, and incomplete. To better serve students, the higher education system needs a data infrastructure that reduces complexity while measuring the outcomes of all students—especially those who have been traditionally underserved.

This brief builds on the 2016 paper series *Envisioning the National Postsecondary Infrastructure in the 21st Century*[2] by exploring a federal student-level data network (SLDN) through both technical and policy lenses.[3] We know policymakers are interested in creating a workable SLDN, as evidenced by the introduction of the bipartisan College Transparency Act of 2017 (CTA) and previous iterations of the Student Right to Know Before You Go Act.[4] This paper, therefore, reinforces the case for building an SLDN while describing the technical, operational, and governance requirements needed to successfully and securely design and implement such a system.[5] Figure ES1, below, shows our key policy recommendations for an SLDN in three primary categories: **Operations & Capacity**, **Data Governance**, and **Privacy & Security**.

The creation of a federal SLDN would transform the higher education system. If higher education stakeholders had access to connected and complete data, and used it thoughtfully, more students would have a true opportunity to succeed in college. Policymakers could make more informed decisions about federal and state investments in higher education. Institutions could benchmark performance and create programs that benefit students. Students and families could choose a best-fit college. And researchers could provide a data-driven foundation for these stakeholders. High-quality data is an essential, yet often overlooked, component of most policy areas.

However, the federal government did not design its systems to collect data to answer all of the questions pertinent to today's student population. For example, the most commonly used graduation rate only measures the percentage of first-time, full-time students who complete their degree or credential at their first institution within six years, leaving out part-time and transfer students. As a result, this rate reflects only about 47 percent of today's students entering college.[6] The federal government collects data at different levels (e.g., institution-level, student-level, loan-level), and its data systems cannot currently connect to each other to provide a clear picture of the complex and varied pathways students take to, and through, higher education—pathways that we now know are highly complex. Given rising college costs, increased interest about return on investment for families and the government, and a lack of clear information about 21st century students' outcomes, we must take action to improve our national data infrastructure and put useful information into the hands of decision makers.

Students, policymakers, and institutions need a limited, core set of consistent, institution- and program- level metrics that serve specific policy, consumer information, and institutional improvement purposes. This brief provides the necessary guidance to design and implement a secure data system that protects student privacy, while equipping key stakeholders with the information they need to make equitable, student success-focused decisions.

Figure ES1: Key Policy Recommendations for a Federal Student-Level Data Network (SLDN)

	RECOMMENDATIONS
Operations and Capacity 	▶ Authorize creation of a federal SLDN. ▶ Leverage existing federal and institutional data to count all students and all outcomes. ▶ Replace components of existing data collections with a federal SLDN. ▶ Shift staffing and data system resources from Integrated Postsecondary Educations Data System (IPEDS) to the SLDN.
Data Governance 	▶ Include key stakeholders on the data governance team to inform data integrity, management, and privacy. ▶ Coordinate on the development and use of unique identifiers to align data across collections. ▶ Adapt best practices from existing data sharing efforts.
Privacy and Security 	▶ Follow data minimization principles to limit data included in the system to only the necessary elements, retain data only as long as needed, and restrict its use to educational purposes. ▶ Host the system in a statistical agency and require adherence to strict privacy and security laws and standards, including conducting routine audits, using encryption technology, and following relevant standards and practices from the National Institute for Standards and Technology, the Fair Information Practice Principles, and other leading protocols. ▶ Implement clear role-based access protocols.

Introduction

Research shows that investing in a college education is worth it.[7] Persistent equity gaps in college access, success, and attainment, however, prove that these investments don't pay off equally for all students at all institutions. We need to know which programs at which institutions provide a return on investment, and for which students. Outcomes vary across institutions, yet the data that students, institutions, and policymakers currently use to make decisions are incomplete and insufficient. Better use of high-quality data will improve student success, especially for underserved students who too often are left out of data metrics and systems.

In the current system, institutions report data—sometimes very similar data—to a variety of entities, including states, accrediting agencies, voluntary data initiatives, and the federal government. Within the federal government, the Integrated Postsecondary Education Data System (IPEDS) includes institution-level, aggregate data on student access, progression, price, and completion at each participating institution. The National Student Loan Data System (NSLDS) is comprised of student- and loan-level data on federal student aid recipients, and the Department of Veterans Affairs (VA) and Department of Defense (DoD) hold administrative data on students receiving veteran education benefits. For the most part, these data systems do not communicate, creating a confusing and disjointed data infrastructure that produces incomplete metrics.

Because data are collected at different levels (e.g., student-level, loan-level, institution-level) and there have been few efforts to match data across systems, our current framework cannot answer critical questions about college enrollment, completion,

costs, and outcomes. Without clear answers about outcomes for today's students, it is difficult for students to make informed choices, for policymakers to promote equitable access to and success in higher education, and for institutions to implement data-driven reforms. Equity and success for all students will remain elusive until policymakers, students, and institutions have the proper information to make decisions.

Over the past ten years, the higher education field has established key metrics to advance student success by measuring postsecondary performance, efficiency, and equity. Using examples from voluntary data collections, IHEP constructed a postsecondary metrics framework that reflects a decade of progress in defining metrics and using data. **Figure 1**, below,

Figure 1: A Field-Driven Metrics Framework

	ACCESS	PROGRESSION	COMPLETION	COST	POST-COLLEGE OUTCOMES
PERFORMANCE	Enrollment	Credit Accumulation Credit Completion Ratio Gateway Course Completion Program of Study Selection Retention Rate Persistence Rate	Transfer Rate Graduation Rate Success Rate Completers	Net Price Unmet Need Cumulative Debt	Employment Rate Median Earnings Loan Repayment and Default Rates Graduate Education Rate Learning Outcomes
EFFICIENCY	Expenditures per Student	Cost for Credits Not Completed Cost for Completing Gateway Courses Change in Revenue from Change in Retention	Time/Credits to Credential Cost of Excess Credits to Credential Completions per Student	Student Share of Cost Expenditures per Completion	Earnings Threshold

Key Student Characteristics

Enrollment Status
Attendance Intensity
Credential-Seeking Status
Program of Study
Academic Preparation
Military Status
Economic Status
Race/Ethnicity
Age
Gender
First-Generation Status

Institutional Characteristics

Sector
Level
Credential/Program Mix
Size
Resources
Selectivity
Diversity
Minority-serving Institution (MSI) Status
Post-traditional Populations
Modality

Source: *Toward Convergence: A Technical Guide for the Postsecondary Metrics Framework:* http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/ihep_toward_convergence_low_2b.pdf

shows the recommended metrics necessary to meet the needs of all relevant stakeholders.[8] This limited, core set of consistent metrics serves specific policy, consumer information, and institutional improvement purposes, while focusing on equity in student access, progression, completion, cost, and post-college outcomes.[9] These metrics would be reported and used at the program- and institution-level, aggregated from the student-level data matched by the system.

Policymakers recognize that improving our national postsecondary data infrastructure is essential. Accordingly, they have shown ongoing interest in improving the quality of information relied upon by both consumers and the policymakers charged with stewarding the federal government's substantial investment in higher education. Legislation like the Student Right to Know Before You Go Acts of 2012, 2013, and 2015, as well as the College Transparency Act of 2017 (CTA), identify a secure, privacy-protected federal student-level data network (SLDN) as the most effective way to correct the duplicative, disconnected, and incomplete nature of the current postsecondary data systems.[10]

Equity and success for all students will remain elusive until policymakers, students, and institutions have the proper information to make decisions.

This brief describes operations and capacity, data governance, and privacy and security recommendations for designing and implementing an SLDN. It also answers basic questions about what a federal SLDN would look like: where it would operate, who would submit data, who would have access to the data, and who would govern the overall system. The data governance team creates rules, processes, and procedures in collaboration with the Commissioner of the National Center for Education Statistics (NCES). Only those with clearance and training would have access to the more granular data—most data users would receive only aggregate institution- or program-level data.

A federal SLDN would streamline the way institutions report data to the federal government, while increasing the quality and usability of the resulting information. Currently, every Title IV-participating institution uses student-level data to calculate aggregate metrics, like graduation rates, for IPEDS. The number of metrics that institutions submit can vary, based on institution type and level. Many institutions must calculate upward of 500 metrics, including, but not limited to, enrollment, completion, and pricing metrics, disaggregated by a variety of subsets of students. By contrast, under a federal SLDN, institutions would instead securely report the student-level data they already hold to the Department of Education (ED). ED would then use the data to calculate and report the aggregate metrics, allowing the SLDN to replace portions of IPEDS. This system would reduce institutional reporting burdens, while allowing ED to calculate even more comprehensive and useful metrics.

For example, under a federal SLDN, student-level data on race/ethnicity would be reported through the system to allow ED to calculate the IPEDS enrollment and graduation rates by race/ethnicity. Without any additional reporting—but informed by field input—ED could disaggregate other measures, such as the IPEDS Outcome Measures, by race/ethnicity. The student-level data would never be shared publicly, but it would provide a more robust underpinning to the public, institution-level metrics. This more complete and efficient system would allow for more comprehensive, equity-focused analyses of aggregate data.

To supplement data submitted by institutions, the SLDN would alleviate duplicative reporting by matching student-level data from institutions to data already collected by the federal government. Agencies like the Department of the Treasury, Social Security Administration, Office of Federal Student Aid (FSA), DoD, and VA would enter data sharing agreements with NCES to produce aggregate reports on post-college workforce outcomes, the cost of a college degree, and student veteran and servicemember access and success. NCES is well-equipped to house this type of system because of its status as an independent, statistical agency and long history managing sensitive data. State and federal policymakers, institutions, and students could then leverage these data when making decisions about investment and improvements in higher education.

More specifically, data feedback loops to institutions and states would provide valuable information on students' pathways to, and through, other institutions, as well as students' post-college outcomes. Colleges and universities want to know how their students fare after leaving their institution—either to pursue subsequent education or to enter the workforce. Right now, however, schools have limited access to information that would answer questions about how prepared their students are to succeed after transfer or graduation. Similarly, states want to understand the development, retention, and flow of human capital to help strengthen their economies and tailor policies to student and workforce needs. Aggregate data or custom queries from an SLDN could answer these types of questions for both states and institutions in more complete and comprehensive ways than existing data.

This paper describes the steps toward an infrastructure design where data are secure, student privacy is protected, and policymakers, institutions, and students all have the aggregate institution- or program-level information they need to make informed decisions. Improvements to the infrastructure would provide an opportunity to count all postsecondary students, all outcomes, and all institutions, while advancing equity through the use of

disaggregated metrics. This paper provides an overview of the technical, governance, and capacity requirements for a federal SLDN, exploring design considerations and policy recommendations in three key categories: **Operations & Capacity**, **Data Governance**, and **Privacy & Security**.

Recommendations for a Federal Student-Level Data Network

While members of the postsecondary policy and advocacy community[11] have underscored the policy imperative and recommended strategies to create a more cohesive data ecosystem, we have not yet fully explored the operational and technical requirements of an improved system.[12] **Figure 2**, below, describes these considerations.

Operations and Capacity

For a data system to function properly, ED needs to assess its internal capacity and staffing to provide the daily resources necessary for operation. Policymakers will need to address funding and internal restructuring in order to implement a federal student-level data network. Recommendations for doing so are laid out in detail below.

1. Authorize the creation of a federal SLDN. The creation of federal SLDN is currently prohibited by federal statute, so Congress needs to act to allow the system to be built. The recently introduced College Transparency Act would overturn the federal prohibition and create a federal SLDN. The Act has the support of over 90 organizations that recognize the need for improved data systems to promote student success.[13] Several institutional associations are included among these supporting organizations, signaling the support of many colleges and universities.

2. Leverage existing federal and institutional data to count all students and all outcomes. A primary goal of a federal SLDN is to count all students and all outcomes to accurately represent the state of today's postsecondary system. Historically, federal collections have only focused on traditional student cohorts (e.g., first time, full time students who enroll in the fall semester, or students who receive federal financial aid).[14] Higher education metrics should also include part-time, transfer, and non-aided students, and measure all pathways and outcomes, such as transfer, completion after transfer, and workforce outcomes.[15]

To calculate these metrics, ED should leverage existing federal and institutional data. If the federal government already holds data, such as for workforce outcomes, federal financial aid receipt, or veteran education benefits, then institutions should not be required to report that information. Instead, institutions should report information not already held by the federal government, such as enrollments, pricing, and completions. This type of information is needed to calculate IPEDS metrics, and will help fill reporting gaps. Institutionally-reported data would be periodically and securely matched with federally-held data to produce useful, aggregate information. For instance, institutions would report student-level enrollment and completion data to NCES, which would combine the data into cohorts. NCES would securely transmit individual-level data, grouped into the pre-determined cohorts, to the Department of the Treasury, which would match the data with earnings information. Treasury would send aggregate workforce information back to NCES on the cohorts, but would not transfer earnings data about individual students. After sending the aggregate data to

Figure 2: Key Policy Recommendations for a Federal Student-Level Data Network (SLDN)

	RECOMMENDATIONS
Operations and Capacity	▶ Authorize creation of a federal SLDN. ▶ Leverage existing federal and institutional data to count all students and all outcomes. ▶ Replace components of existing data collections with a federal SLDN. ▶ Shift staffing and data system resources from Integrated Postsecondary Educations Data System (IPEDS) to the SLDN.
Data Governance	▶ Include key stakeholders on the data governance team to inform data integrity, management, and privacy. ▶ Coordinate on the development and use of unique identifiers to align data across collections. ▶ Adapt best practices from existing data sharing efforts.
Privacy and Security	▶ Follow data minimization principles to limit data included in the system to only the necessary elements, retain data only as long as needed, and restrict its use to educational purposes. ▶ Host the system in a statistical agency and require adherence to strict privacy and security laws and standards, including conducting routine audits, using encryption technology, and following relevant standards and practices from the National Institute for Standards and Technology, the Fair Information Practice Principles, and other leading protocols. ▶ Implement clear role-based access protocols.

NCES, Treasury should destroy the data post-matching by following NIST guidelines for media sanitization.[16]

3. Replace components of existing data collections with a federal SLDN. Large portions of IPEDS, such as the Fall Enrollment, 12-Month Enrollment, Graduation Rates, 200% Graduation Rates, Outcome Measures, Completions, Student Financial Aid, and portions of the Admissions surveys, could be replaced by a federal SLDN. The governance team, data experts, and federal agencies should examine existing data collections and replace components that would be duplicated by a student-level system.

4. Shift staffing and data system resources from IPEDS to the SLDN. The SLDN will require staff, software, hardware, training resources, and mechanisms for engaging with the governance team. To ensure comprehensive data governance and compliance with privacy and security standards, funding and resources should shift from the IPEDS collection to the SLDN, as the system will fulfill the student components of IPEDS reporting. ED will need to make internal capacity adjustments to phase out old collection mechanisms, phase in new systems, conduct trainings, respond to inquiries from reporting institutions and states during the transition, and convene a data governance team to assure the success of the system.

Data Governance

Data governance is a systemic process that formalizes how people interact—or do not interact—with data systems.[17] This process requires technical input and expertise, and should be an integral facet of the data infrastructure strategy. ED should form a data governance team early in the planning phase to manage the integration of complex education and workforce data systems for the federal SLDN. Statute should require that this team involve representation from each of the agencies involved in the network, including leaders like chief information officers, privacy officers, and database administrators; experts in privacy and security, postsecondary data quality, and consumer protection; governmental and non-governmental researchers; institution, state, and agency representatives; and data users, like students and policymakers. With diverse representation of subject matter and policy experts, the governance team will be able to develop policies and processes that create a sustainable and flexible data network.

Once assembled, the data governance team would be responsible for the research and execution of memorandums of understanding (MOUs) and data sharing agreements between multiple stakeholders. The team would also be responsible for establishing a comprehensive governance program that ensures security, privacy, confidentiality, integrity, and appropriate accessibility of the data, including the development of a data dictionary. To accomplish this non-exhaustive list of tasks for the governance team, policymakers should consider the following recommendations.

1. Include key stakeholders on the data governance team to inform data integrity, management, and privacy. An SLDN's governance team should require engagement with the key stakeholders and prioritize, first and foremost, the needs of students. The governance team should include representatives of students and families, consumer protection advocates, institutions, states, privacy and security experts, postsecondary data experts, and data users, such as policy analysts and researchers. This team should guide system design and implementation by defining all aspects of data management and use. By leveraging the expertise and experience of these groups, the data governance team should work with developers to create a system that meets statutory requirements, agency parameters, privacy and security protocols, and consumer information needs. The governance team should also convene privacy and security experts to regularly review and audit data and technology standards. The team should also create data access policies, including guidelines for developing a public tool (similar to NCES PowerStats) and restricted use licenses for researchers, as well as data minimization, retention, destruction, and breach protocols.

2. Coordinate on the development and use of unique identifiers to align data across collections. A unique identifier for student-level data plays the pivotal role of allowing these data to be matched across collections. While the data matching process comes with the aforementioned privacy and security considerations, accurate matches require an identifier or a series of identifiers. A federal data network can learn from states and other data initiatives that have securely matched data across sources. For instance, the Western Interstate Commission for Higher Education's (WICHE) Multistate Longitudinal Data Exchange (MLDE) leverages the National Student Clearinghouse's (NSC) proprietary identity resolution process to match persons between education and workforce data sources, like state departments of education, labor market information offices, and NSC.[18] Similarly, the Arkansas Research Center utilizes a dual-database architecture to match records while protecting students' identities. Under this approach, personally identifiable information (PII) is secured in one database and matched with a temporary unique, random identifier to align with other databases.[19] Federal agencies can use these and other models to create data sharing agreements and processes that protect student privacy and secure PII as carefully as possible.

3. Adapt best practices from existing data sharing efforts. Despite the lack of a comprehensive, national data network,

some states, institutions, and federal agencies have undertaken efforts to match data as part of their decision-making processes. Lessons learned from these initiatives in metric alignment, capacity requirements, and MOUs can inform the changes needed to implement a federal SLDN. For example:

- The College Scorecard required the coordination of large administrative datasets from within ED and the U.S. Department of the Treasury. Treasury matched student-level data from FSA with earnings data from administrative tax records for federally-aided students and shared aggregate, institution-level results with ED to match institution-level IPEDS and FSA data.[20] A national SLDN would build similar matches to allow workforce outcome metrics to be calculated for aided and non-aided students. ED and Treasury can leverage lessons learned and MOUs from this initial data sharing to implement the improved data network.

- Some states are participating in data sharing initiatives to understand student outcomes while accounting for student mobility. WICHE's MLDE connects four state data systems to measure wage and employment outcomes across states. This initiative provides valuable insight into challenges for data sharing that arise when different agencies work together to fill gaps in knowledge.[21] The data governance team should use the lessons learned in establishing complex MOUs and navigating state legislation on student data privacy and security to inform its process.[22]

- The University of Texas (UT) System partnered in fall 2016 with the U.S. Census Bureau to connect education data with salary and jobs data from the Longitudinal Employer-Household Dynamics (LEHD) program.[23] This partnership includes a 10-year agreement and provides both Census and UT with data that connect degree attainment with labor market outcomes. These data augment UT's abilities to understand the impact of higher education on their students who stay in-state, as well as those who move out of Texas, and is intended to inform policy decisions on student debt, program assessment, and student advising initiatives.

- To provide veterans with better information about postsecondary opportunities, the Department of Labor, DoD, ED, and VA have experimented with collaboration and data sharing.[24] These are some of the primary agencies that would need to coordinate to implement a comprehensive system that includes all student- and policy-relevant metrics, especially to meet the needs of veterans and simplify the reporting on data about service members' and veterans' use of financial aid.

Privacy and Security

The value proposition for a federal SLDN is clear: it will close gaps in current data systems and, in turn, will answer stakeholder questions about access, success, and outcomes for all students while providing the information necessary to protect taxpayer investment in higher education. With answers to these questions, students will be able to make more informed decisions, policymakers will be able to drive changes in policy and practice to protect both students and taxpayers, and institutions will be able to tackle continuous improvement efforts. All of this will result in more equitable and improved student outcomes. New data matches will require thoughtful and ongoing attention to privacy and security, but federal agencies will be able to mitigate risk by leveraging industry best practices and complying with relevant federal information technology (IT) system requirements, including the Federal Trade Commission's Fair Information Practice Principles (FIPPs) and the standards set by the National Institutes of Standards and Technology (NIST).[25] Protecting the privacy and security of student data must remain the top priority for those who report, collect, or aggregate the data. They must ensure that the SLDN adheres to all laws governing data collection and use, including those listed in **Figure 3**, on the next page.

Continuously updating and properly executing privacy and security protocols will protect student data from misuse and unauthorized access. The data governance team should routinely revisit the system's privacy and security protocols, and consult with experts to make improvements. In addition, the data network itself should be audited on a regular basis, and encryption technology should be used to protect and secure the data. As technology to protect and secure data improves, so should the SLDN. Furthermore, data in an SLDN should not be used to punish or take corrective action against students or for corporate gain. The following recommendations detail ways to protect student data while implementing a more robust data network.

Protecting the privacy and security of student data must remain the top priority for those who report, collect, or aggregate the data.

1. Follow data minimization principles to limit data included in the system to only the necessary elements, retain data only as long as needed, and restrict its use to educational purposes. Data minimization is a key privacy principle. A federal SLDN should contain only the data elements that are needed to answer questions of national importance about college access, success, cost, and outcomes. The data should be restricted to educational purposes and retained only as long as necessary. The data governance team should determine which elements to include in the network to serve the purposes outlined in statute like consumer information, policy-

Figure 3: Laws and Agencies Related to Student Data Privacy and Security

LAWS AND PRINCIPLES	CONSIDERATION
The Family Educational Rights and Privacy Act of 1974 (FERPA)	FERPA protects the access and use of student data by all educational agencies and institutions that receive federal funding. Once a student attends a postsecondary institution, the rights formerly provided to the parent are transferred to the student. FERPA also supports the protection of personally identifiable information (PII).
The Privacy Act of 1974	The Privacy Act of 1974 protects the privacy of records created and used by the federal government. Accordingly, it would apply to all data stored within a federal SLDN. In addition to stating the rules that the government must follow when collecting data about a person, the law ensures that the government cannot disclose data about a person without that person's permission unless the disclosure meets one of the 12 broad statutory exemptions outlined in the statute.
Confidential Information Protection and Statistical Efficiency Act (CIPSEA)	CIPSEA provides strong confidentiality protection to data used for statistical purposes. It protects the data from law enforcement, taxation, and regulatory use.
Higher Education Act (HEA)	HEA authorizes numerous federal aid programs that provide support to individuals and institutions. HEA provisions govern the permissible use of data collected through the Free Application for Federal Student Aid (FAFSA) and in the National Student Loan Data System (NSLDS).
The Federal Information Security Management Act of 2002 (FISMA)	FISMA applies to federal IT systems and other IT systems that hold federal data. It requires those systems to adhere to common national standards regarding information security protection and to utilize a risk-based approach when securing data. Annual reviews through the Office of Management and Budget are also required to ensure the security program in place is adequate.
E-Government Act of 2002	The E-Government Act of 2002 protects how data are collected, stored, and used in a federally-held IT system. The law requires a privacy impact assessment to be completed, as well as the posting of privacy notices regarding a system's data collection practices.
The Gramm-Leach-Bliley Act (GLBA)	The GLBA impacts how institutions collect, store, and use student financial records containing PII.
The Fair and Accurate Credit Transaction Act of 2003 (FACTA)	FACTA also impacts how institutions collect, store, and use student financial records containing PII.
Fair Information Practice Principles (FIPPs)	The Federal Trade Commission's FIPPs are guidelines for how entities collect and use personal information, and the safeguards they use to assure adequate privacy protection. These principles are part of the Privacy Act of 1974 (see above).
AGENCY/OFFICE	**RESPONSIBILITY**
National Institute of Standards and Technology (NIST)	This agency resides within the U.S. Department of Commerce and develops publications and standards for privacy and security as part of its statutory responsibility under FISMA.
Office of Inspector General (OIG) at ED	ED's Office of Inspector General promotes efficiency, effectiveness, and integrity of ED operations and programs. OIG conducts the annual FISMA compliance audit for ED.[26]
Privacy Technical Assistance Center (PTAC)	This center resides within ED and serves as a "one-stop" resource for education stakeholders to learn about data privacy, confidentiality, and security practices related to student-level longitudinal data systems and other uses of student data. PTAC also provides training materials and opportunities to receive direct assistance in the aforementioned topics.[27]

Source: Selected sections of table are excerpted from *Understanding Information Security and Privacy in Postsecondary Education Data Systems (Grama, 2016)*. http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/information_security_and_privacy.pdf

making, and institutional improvement. A proposed limited list of data elements is included in the appendix to this paper. A federal data system should not include sensitive data elements on students or their families, including, but not limited to, those related to health, social-emotional, immigration, or disciplinary records. Sensitive elements such as these should be statutorily prohibited from being collected.

2. Host the system in a statistical agency, and require adherence to strict, industry-leading privacy and security standards, including conducting routine audits, using encryption technology, and following relevant standards and practices from NIST, FIPPs, and other leading protocols. ED includes several offices that collect, process, and analyze data at different levels. The NCES, the proposed agency to house the federal SLDN per the College Transparency Act, adheres to statistical agency standards and laws and has maintained secure databases without incident for decades. As part of ongoing system audits and reevaluations, ED leadership should review and enhance privacy and security protocols in accordance with applicable legislation, like FISMA and CIPSEA, and emerging best practices in the field, including encryption technologies. Additionally, agencies involved in data sharing should leverage information security publications from NIST to ensure that they maintain the privacy of all students while keeping their data secure and protected from internal and external breaches.

To ensure that ED and other participating federal agencies safeguard student data, policymakers should require agencies to adhere to strict privacy and security standards for data matching, use, and storage, including applying penalties for willful violations. Such a requirement would signal the seriousness of privacy and security standards at the federal level. Any applicable statute, however, must permit standards to evolve with developing technologies and emerging threats. Statutory language, therefore, should not be prohibitively explicit. For example, policy should require the data network to comply with the relevant portions of federal and industry-leading standards, such as those delineated by NIST, FISMA, and FIPPs.

3. Implement clear role-based access protocols. Role-based access to data ensures that the right people will have access to the right data. The data governance team is responsible for creating access guidelines for the system. The team could suggest using technology like multifactor authentication to safeguard data from improper access and use.

How Do We Get There?

With a framework for which metrics to collect (see **Figure 1**), the federal government needs to reconstruct the national data infrastructure to facilitate the collection and use of these data. Policymakers and federal agencies should collaborate to improve data matching and utilize the data they already have for decision-making. These matches, however, will take time to implement. Creating and maintaining a federal SLDN will involve five phases (see **Figure 4**, below):

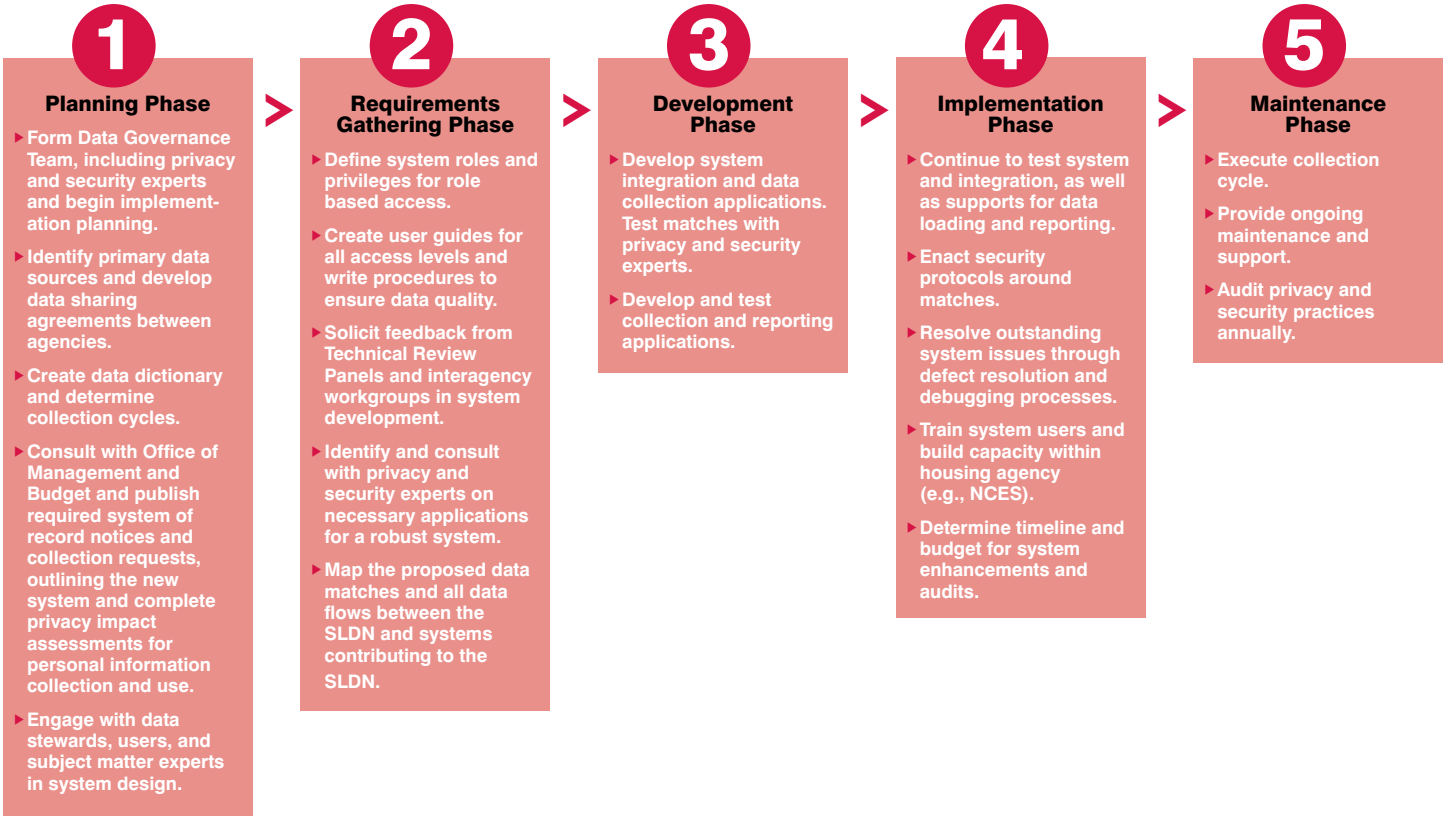
1. In the **planning phase**, the governance team will create the procedural foundation for the system. They will outline the processes needed to implement the system, including the data sharing agreements required to collect and use data. The governance team will use the planning phase to ensure that the system meets the statutory requirements in the Higher Education Act for federal data collection. The team will also confirm data sharing agreements across participating federal agencies. As a result, this could be the most time-consuming and most politically complicated of the five project phases.

2. The **requirements gathering phase** will lay the technical groundwork to build a secure system by defining system roles and privileges for users, as well as creating guides for users at all access levels. During this time the data governance team will actively gather the input of key stakeholders on the design and uses of the system by convening interagency workgroups and review panels.

3. In the **development phase**, system architects will build the system and test it with end users and data providers, such as institutions and federal agencies. Architects will construct how the data systems across agencies will match and integrate data. They will also determine which mechanisms are best for data collection and reporting, and then test these assumptions with the data governance team and workgroups.

4. The **implementation phase** will include one year of trouble-shooting with end users and data providers to solve technical issues around data input, reporting, and management. ED will also spend this year building staff capacity and training them on

Figure 4: Phases for Creating and Implementing a Student-Level Data Network



the intricacies of the system. This phase readies the system for use. At the conclusion of the phase, ED will deploy the system.

5. The **maintenance phase** will continue indefinitely. It initiates data collection and includes processes for support and system improvements, including the refinement of data elements in the system.

Conclusion

Better data are a necessary tool for helping more students, especially low-income students and students of color, succeed in higher education. Upgrades to the national postsecondary data infrastructure are needed to answer questions that will allow for improved student outcomes through policymaker, practitioner, and student action. A student-level data network will bring the postsecondary data infrastructure closer to the ideal, where state and federal policymakers, students and families, and institutions all have the information they need to make important decisions and close equity gaps. The creation of an SLDN will require significant federal action and inter-agency collaboration. Nonetheless, a failure to do so will hamper broad, bipartisan efforts to enhance student choice, transparency, and improved student outcomes. Students can wait no longer for the information and transparency that a more efficient and effective data system can provide.

Endnotes

1 Ma, J., Pender, M., & Welch, M. (2016). Education Pays 2016. The College Board. Retrieved from: https://trends.collegeboard.org/sites/default/files/education-pays-2016-full-report.pdf

2 Institute for Higher Education Policy. *Envisioning the national postsecondary infrastructure in the 21st century*. Retrieved from Institute for Higher Education Policy website: http://www.ihep.org/postsecdata/mapping-data-landscape/national-postsecondary-data-infrastructure

3 A federal student-level data network (SLDN) aggregates student-level data reported by institutions and federal agencies to create aggregate metrics for decision-making and policy-making.

4 Roberson, A.J., & Voight, M. (2017). *Better data on the horizon: An analysis of evolving student-level data legislation*. Retrieved from Institute for Higher Education Policy website: http://www.ihep.org/postsecdata/resources/better-data-horizon-analysis-evolving-student-level-data-legislation-0

5 College Transparency Act of 2017, H.R. 2434, 115th Congress (2017-2018). Retrieved from: https://www.congress.gov/bill/115th-congress/house-bill/2434/text

6 IHEP analysis of IPEDS 2015 data.

7 Ma, J., Pender, M., & Welch, M. (2016). Education Pays 2016. The College Board. Retrieved from: https://trends.collegeboard.org/sites/default/files/education-pays-2016-full-report.pdf

8 For more information, see Janice, A., & Voight, M. (2016). *Toward convergence: A technical guide for the Postsecondary Metrics Framework*. Retrieved from Institute for Higher Education Policy website: http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/ihep_toward_convergence_low_2b.pdf

9 A list of underlying data elements can be found in the Appendix.

10 Student Right to Know Before You Go Act of 2012, S. 2098, 112th Congress (2011-2012). Retrieved from: https://www.congress.gov/bill/112th-congress/senate-bill/2098; Student Right to Know Before You Go Act of 2013, H.R. 1937, 113th Congress (2013-2014). Retrieved from: https://www.congress.gov/bill/113th-congress/house-bill/1937/text; Student Right to Know Before You Go Act of 2015, H.R. 2518, 114th Congress (2015-2016). Retrieved from: https://www.congress.gov/bill/114th-congress/house-bill/2518/text; College Transparency Act of 2017, H.R. 2434, 115th Congress (2017-2018). Retrieved from: https://www.congress.gov/bill/115th-congress/house-bill/2434/text

11 The Postsecondary Data Collaborative (PostsecData) is coalition of over 35 organizations dedicated to the use of high quality data to improve student success and close equity gaps. For more information, see: www.ihep.org/postsecdata

12 For more information, see Institute for Higher Education Policy. *Envisioning the national postsecondary infrastructure in the 21st century*. Retrieved from Institute for Higher Education Policy website: http://www.ihep.org/postsecdata/mapping-data-landscape/national-postsecondary-data-infrastructure

13 Institute for Higher Education Policy. Postsecondary data collaborative and Workforce Data Quality Campaign applaud bipartisan, bicameral College Transparency Act. (2017, May 23). Retrieved from Institute for Higher Education Policy website: http://www.ihep.org/press/opinions-and-statements/postsecondary-data-collaborative-and-workforce-data-quality-campaign

14 Starting in 2017, IPEDS released data from its Outcome Measures (OM) component, including completion data for non-first-time and non-full-time students, six and eight years after initial enrollment.

15 For recommended metrics and definitions, see Janice, A., & Voight, M. (2016). *Toward convergence: A technical guide for the Postsecondary Metrics Framework*. Retrieved from Institute for Higher Education Policy website: http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/ihep_toward_convergence_low_2b.pdf

16 The federal government used this type of data sharing process to produce earnings metrics on the College Scorecard, but instead of using institutionally reported data on all students, it used student-level NSLDS data only on federal aid recipients. See: https://collegescorecard.ed.gov/data/documentation/

17 Hosch, B. (2017, May 31). *Beyond data governance to data strategy*. AIR Annual Forum, Washington D.C. Retrieved from: http://www.stonybrook.edu/commcms/irpe/reports/_presentations/DataStrategy_Hosch_2017_05_31.pptx

18 Prescott, B. (2014 July). *Beyond borders: Understanding the development and mobility of human capital in an age of data-driven accountability*. Retrieved from: http://www.wiche.edu/info/longitudinalDataExchange/publications/MLDE_BeyondBorders.pdf

19 Arkansas Research Center. Dual Database Architecture. Retrieved from: https://arc.arkansas.gov/arcweb/?page_id=2

20 U.S. Department of Education. (2015). *Using federal data to measure and improve the performance of U.S. institutions of higher education*. Retrieved from: https://collegescorecard.ed.gov/assets/UsingFederalDataToMeasureAndImprovePerformance.pdf

21 Prescott, B., & Lane, P. (2016). *Fostering state-to-state data exchanges*. Retrieved from Institute for Higher Education Policy website: http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/fostering_state-to-state_data_exchanges.pdf

22 Data Quality Campaign. (2016). Student data privacy legislation: A summary of 2016 state legislation. Retrieved from: http://dataqualitycampaign.org/resource/2016-student-data-privacy-legislation/; Data Quality Campaign. (2015). Student data privacy legislation: What happened in 2015 and what is next? Retrieved from: http://dataqualitycampaign.org/resource/student-data-privacy-legislation-happened-2015-next/

23 University of Texas System (2016, September 22). UT System partners with U.S. Census Bureau to provide salary and jobs data of UT graduates across the nation. Retrieved from: https://www.utsystem.edu/news/2016/09/22/ut-system-partners-us-census-bureau-provide-salary-and-jobs-data-ut-graduates-across

24 The White House (2016, November 11). FACT SHEET: Ensuring veterans have the tools to succeed. Retrieved from: https://www.whitehouse.gov/the-press-office/2016/11/11/fact-sheet-ensuring-veterans-have-tools-succeed

25 Federal Trade Commission. Fair information practice principles. Retrieved from: https://web.archive.org/web/20090331134113/http://www.ftc.gov/reports/privacy3/fairinfo.shtm

26 U.S. Department of Education. Office of Inspector General. Retrieved from: https://ed.gov/about/offices/list/oig/index.html

27 U.S. Department of Education, Privacy Technical Assistance Center. About PTAC. Retrieved from: http://ptac.ed.gov/About

Appendix: Elements to Include in a Federal Student-Level Data Network

DATA ELEMENT	METRIC SUPPORTED	WHO REPORTS?/ DATA SOURCE	REQUIRED FOR IPEDS REPORTING
Social Security Number	Record matching	All linked sources	n/a
Unique student ID	Record matching	Created for system operation	n/a
Name	Record matching	Institution	n/a
Date of birth	Age at time of entry; Record matching	Institution	Y
State of residency	IPEDS enrollment; Cost of Attendance	Institution	Y
Race/ethnicity	Race/ethnicity disaggregate	Institution	Y
Gender	Gender disaggregate	Institution	Y
Family Income	Economic status disaggregate	FSA (for aided); Institution (for non-aided)	Y
Military/Veteran Status	Military status disaggregate	Departments of Defense and Veterans Affairs	N
First generation college student status	First generation disaggregate	FSA and Institution	N
Institution ID	Record matching; Calculation of institution-level metrics	IPEDS/Office of Postsecondary Education at ED	Y
Date first enrolled	IPEDS yield rate; Cohort determination; Time to Credential; Enrollment	Institution	Y
Enrollment by term	Retention; Persistence	Institution	Y
Credential-seeking status	Cohort determination; Disaggregate; Graduate Education Rate	Institution	Y
Transfer Status	Enrollment status	Institution	N
Program of Study (CIP Code)	Program of study selection; Disaggregates	Institution	Y
Enrollment mode	Modality	Institution	Y
College-ready status	Academic preparation disaggregate	Institution	N
Number of credits attempted	Credit accumulation; Credit completion ratio; Credits to degree; Attendance intensity (Full-time, part-time)	Institution	N
Number of credits completed	Credit accumulation; Credit completion ratio; Credits to degree; Attendance intensity (Full-time, part-time)	Institution	N
Completion of math gateway course in first year	Gateway Course Completion	Institution	N
Completion of English gateway course in first year	Gateway Course Completion	Institution	N
Date of credential award	Graduation rate; Completers; Time to credential; IPEDS Degrees conferred	Institution	Y
Credential award level	IPEDS degrees conferred; Disaggregate	Institution	Y
CIP Code of credential award	IPEDS degrees conferred; Disaggregate	Institution	Y
FAFSA submission flag	Financial aid application	FSA	N
EFC	Net Price; Unmet Need	FSA	N
Tuition and fees	Cost of attendance; Net Price; Unmet Need	Institution	Y
In-state (or in-district) tuition eligibility	Cost of attendance; Net Price; Unmet Need	Institution	Y
Room and board	Cost of attendance; Net Price; Unmet Need	Institution	Y
Living arrangement	Cost of attendance; Net Price; Unmet Need	Institution	Y
Books and supplies	Cost of attendance; Net Price; Unmet Need	Institution	Y
Other expenses	Cost of attendance; Net Price; Unmet Need	Institution	Y
Grant amounts, by source (federal, state, institutional) and type (need/non-need-based)	IPEDS grant awards; Net Price; Pell receipt/Economic Status disaggregate	FSA and Institution	Y
Loan amounts, by source (federal, state, institutional, private), type (Subsidized Stafford, Unsubsidized Stafford, Perkins, PLUS), and interest rate	IPEDS loan amounts	FSA and Institution	Y
Work-study amounts (by source)	Net Price; Unmet Need	FSA and Institution	N
Military benefits amount (by source and type)	GI Bill awards; Tuition Assistance Program awards	Departments of Defense and Veterans Affairs	Y
Cumulative debt	Cumulative Debt; Repayment rate	FSA	N
Date entering repayment	Repayment rate	FSA	N
Loan payment amount	Repayment rate	FSA	N
Remaining debt	Repayment rate	FSA	N
Repayment status	Cohort default rate; Repayment rate	FSA	N
Annual earnings	Employment rate; Earnings; Earnings threshold	Department of Treasury	N

INSTITUTE FOR HIGHER EDUCATION POLICY
1825 K Street, NW, Suite 720
Washington, DC 20006

202 861 8223 TELEPHONE
202 861 9307 FACSIMILE
www.ihep.org WEB

The Institute for Higher Education Policy (IHEP) is a nonpartisan, nonprofit organization committed to promoting access to and success in higher education for all students. Based in Washington, D.C., IHEP develops innovative policy- and practice-oriented research to guide policymakers and education leaders, who develop high-impact policies that will address our nation's most pressing education challenges.