Building a Student-Level Data System

BEN MILLER

CENTER FOR AMERICAN PROGRESS

MAY 2016

Center for American Progress

POSTSEC DATA

Ben Miller is senior director for postsecondary education at the Center for American Progress.

Executive Summary

Introduction

In late 2015, the Obama administration unveiled a revised College Scorecard—a massive release of federally held higher education data.[1] Comprising more than 1,700 indicators for over 7,000 colleges, the College Scorecard showed the power of unlocking federal administrative data.[2] It included the most comprehensive picture to date of student loan repayment rates, the earnings of students receiving federal aid, and debt levels at the institutional level. It also included data for multiple cohorts of students and in most cases disaggregated the data in several different ways, including by students' gender, income upon entering college, and dependency status. Most of these data had never previously been available to the public.

Yet for all their impressiveness, the data also had holes. Both the earnings and completion data only had information on students who received federal student aid. In the worst instances, this meant results represented a minority of the enrollment at some colleges, excluding thousands of students. This led to the possibility of significantly underrepresenting completion rates or the extent to which those institutions provide economic return to their students.

The U.S. Department of Education (ED) has no way to fix these problems under current federal law. This is due to a provision added to the Higher Education Act in 2008 that prohibits ED from creating a new database of information that collects individual-level data on all students enrolled in higher education—referred to hereafter as a student-level system. More commonly known as the "unit record ban," this provision only allows ED to operate student-level databases that are necessary for the operation of the federal student aid programs and that existed prior to 2008.[3]

This paper argues that Congress should overturn that ban and allow the creation of a student-level system at the federal level.

The Benefits of a Student-Level Data System

Ending the unit record ban would provide substantial benefits for institutions, students and families, states, the federal government, the public and others. It would finally make it possible to answer important questions about the effectiveness, efficiency, and impact of America's postsecondary education system. To put it another way, the U.S. Department of Education hands out roughly $130 billion in federal financial aid each year; ending the unit record ban would let everyone involved in the system finally know what they are getting for their money.[4]

For institutions, a student-level data system would also be a powerful tool that would allow them to better document success, gain new information on outcomes, assist in efforts to advocate for additional funding, and potentially help with recruitment. By linking to other federal databases, such as those that contain workforce information, a student-level data system would also grant institutions access to outcomes information that they currently have significant difficulty obtaining. Finally, a student-level data system could help institutions more easily fulfill existing federal reporting requirements, potentially saving over a half a million hours a year.[5]

Meanwhile, for students and families, a student-level system offers access to powerful new sources of data that could help them make better choices about college. Students would be able to generate more personalized information by program or institution about completion, repayment, and earnings outcomes for people similar to them. These data would close some of the significant information gaps that exist now, helping families make better judgments about the proper amounts to pay or borrow for college. It would also help to exert greater market pressure on colleges by strengthening consumers' capacities to vote with their feet and their dollars.

States, too, would benefit from a student-level system. Such a system would help states understand what happens to students who leave their borders. States could also better evaluate the effectiveness of their own aid programs and understand the level of indebtedness of students in their state.

Finally, the student-level data system will also help the federal government, policymakers, and the public better assess the state of American higher education. The federal government would be able to compare the results for students receiving federal aid versus those who do not, allowing it to understand if federal programs achieve desired results and to identify potential policy changes. A federal student-level data system would also benefit the government from a societal standpoint, allowing it to track higher education's effectiveness in promoting social mobility and economic growth by looking at the status of individuals before and after college.

These benefits are not possible with existing federal data collections. For example, ED already knows a lot about individual students who receive federal financial aid, but without data on students who did not receive aid, it cannot contextualize the results it sees. Meanwhile, federal data collections that are

not limited to students receiving federal aid have other problems. For example, institutions report a wealth of data to the National Center for Education Statistics (NCES) about graduation rates and pricing. Yet these rates are calculated in a way that does not accurately capture results for the increasingly large numbers of students who transfer or attend part-time. And while the federal government can address some of these challenges at a national level through large sample surveys, these collections are not frequent enough to provide up-to-date pictures of outcomes, and they are unable to report results at the institutional level.

Unlike other ways of improving our national postsecondary data infrastructure, the federal student-level system does not currently exist. Nor can it exist until Congress repeals the ban prohibiting its creation.

The reasons driving opposition to the student-level system are varied. This report attempts to assuage these concerns by laying out a concrete, coherent vision for how a student-level system could be structured to provide the necessary data for all segments of the postsecondary system while also minimizing added burden to institutions, protecting security and privacy, and leveraging existing operational processes.

The Main Functions of a Student-Level Data System

This paper envisions a student-level system that strengthens ties between today's unconnected state, institutional, and federal databases. The result would be a system that allows for better data analysis while simultaneously lowering the information asymmetry students currently face. To accomplish these goals, a successful system would focus on six main functions:

1. **Construct a complete picture of all students enrolled in college.** This forms the necessary base data for better tracking outcomes.
2. **Collect information on non-federal financial aid and pricing.** These data are necessary to establish better estimates of students' unmet financial need and their need and any received aid contribute to their outcomes. Access to this information would allow institutions and states to evaluate the effectiveness of their aid programs as well as enable the federal government to examine the impact of their programs in relation to institutional and state policies and practices.
3. **Disaggregate data by major student subgroups.** Breaking down data to reflect outcomes for students based upon their age, economic status, and other characteristics not currently available in national data sets can help to establish more personalized information for guiding potential students and also aid states and institutions in identifying and remedying gaps in outcomes.

4. **Link to other federal data sources.** In order for a student-level system to be worthwhile for states and institutions, it must be able to link to other federal data that these parties could not otherwise obtain. This includes information on student earnings after college, where former students end up enrolling, and long-term outcomes for federal student aid.
5. **Report data back to colleges and states.** Simply unlocking federal data is not enough. States and institutions must be able to obtain useful information about their students.
6. **Make aggregated data public.** Students, policymakers, and others must also benefit from this system by being able to obtain aggregated data that (a) help potential consumers know how people with similar backgrounds have fared in the past; (b) allow for the evaluation of broad trends throughout America's postsecondary system; and (c) judge outcomes by institution and programs within a given college for transparency and accountability purposes.

A review of existing federal data collections reveals that moving to such a system would not entail massive changes. For one, institutions already report much of the data needed to calculate completion and transfer rates to ED for the purposes of operating the federal student aid programs, but only for aided students. In addition, many colleges are already reporting these same data on enrollment, completion, and other indicators for all students to a private third-party vendor known as the National Student Clearinghouse (NSC). For institutions that use the NSC, moving to a student-level system would simply mean passing along information they are already putting together.

To maximize the benefits of a student-level system, this paper recommends asking colleges to report a few more indicators. These are in addition to data items already collected, such as a student's gender, age, federal financial aid data, and other indicators. These include:

▶ Student characteristics: race/ethnicity, military status, and level in college
▶ Student attainment progress: a flag for college readiness and information on credits attempted and earned in a given term
▶ Data for net price: amount charged to each student before subtracting any grant aid, state grant aid, institutional grant aid, institutional loans, and non-federal, non-institutional loans

In all cases, this report recommends adding these additional indicators through the same reporting process that colleges currently carry out for federal financial aid purposes. This reduces the likelihood of colleges needing to report data twice on the same students.

Four Options for Building a Student-Level System

This report explores building a student-level system through four approaches:

1. As an expansion of existing systems within the Office of Federal Student Aid
2. As a new database within the Office of Federal Student Aid
3. As a new data collection within the National Center for Education Statistics (NCES), which is housed within the Institute of Education Sciences
4. A hybrid option in which the data are stored in existing FSA systems but NCES oversees public data dissemination and transparency.

While any of the options outlined above are viable, this report ultimately recommends the hybrid option as the best approach. Doing so combines the best of both worlds. FSA has by far the most resources within ED as well as special contracting flexibility that allows it to more easily attract high-quality vendors. It also needs to make sure the student-level system can fit within existing operational needs to avoid complications for schools. FSA, however, does not have significant experience with establishing useful and accessible tools for the public to access aggregated data or for researchers to access de-identified student-level data through a rigorous approval process. Therefore, allowing NCES to receive an extract of this system with the expectation that it will handle public release and researcher access (tasks that are part of its mission already) allows FSA to focus more on its core operational functions.

Recommendations

The key recommendation for this report is an obvious one: end the ban on creating a federal student-level system. Until Congress acts to do so, however, there are intermediate steps available that would make existing federal collections more useful.

Add data dissemination and disclosure to Federal Student Aid's organizational goals. FSA already sits on a wealth of information that would be useful for institutions, states, students, and policymakers; the release of the College Scorecard is testament to what just scratching the surface can bring to light. Nevertheless, the office needs greater incentives to explore data dissemination and transparency. To do so, Congress should adjust the goals and purposes of FSA to more clearly include transparency and data release. This should also become a key objective tied to the bonus paid to the chief operating officer of FSA.

Modernize the National Student Loan Data System. One major challenge with unlocking FSA data is that the main system it uses—the National Student Loan Data System—is clunky and outdated. This system should be modernized by building it using a more current programming language and holding a new competition that draws on larger companies with more resources to administer it.

Create a researcher license for FSA Data. The sheer amount of taxpayer money spent in the federal aid programs demands more transparency. To help with this challenge, FSA should establish a special researcher license that allows approved individuals to gain access to de-identified student records. Such a process could follow the model used by the Internal Revenue Service that recently allowed economists Raj Chetty and Emmanuel Saez to conduct a groundbreaking study using tax records.[6]

Convene technical review panels for necessary new indicators. Many of the new indicators outlined above would be useful for the federal government to collect, even if it is not through a comprehensive student-level system. To make that possible, as well as prepare for the possibility of Congress overturning the ban, NCES should start holding technical review panels to define indicators such as a student's college readiness, credits attempted and earned, and others.

Improve alignment between institutional research and financial aid reporting. This paper strongly recommends establishing the student-level data system through an expansion of the existing individual-level reporting that already occurs for the purposes of the federal financial aid programs. This presents the best path for balancing the goals of comprehensive data coverage while minimizing additional work for institutions.

Doing so, however, could require some changes for institutions. That is because this system would start to calculate outcomes from transactional data that either registrars or the financial aid office currently report. At most colleges, however, these offices do not typically generate data sent to NCES; the office of institutional research usually plays this role. This matters because the two offices may have different protocols for cleaning up data, with the institutional research office having more experience in this area. To better adjust to such a system, institutions should rethink the structures and roles of institutional research offices to order to allow them to play a greater role in overseeing and verifying the accuracy of transactional data.

Building a Student-Level Data System

Introduction

Early on the morning of September 12, 2015, the Obama Administration unveiled what is arguably the largest release of higher education data ever as part of a revamped College Scorecard.[7] Comprising more than 1,700 indicators for over 7,000 colleges, the College Scorecard showed the power of unlocking federal administrative data.[8] It included the most comprehensive picture to date of student loan repayment rates, the earnings of students receiving federal aid, and debt levels at the institutional level. It also included data for multiple cohorts of students, and in most cases disaggregated the data in several different ways, such as by students' gender, income upon entering college, and dependency status.

Yet for all their impressiveness, the data also had substantial holes. Both the earnings and completion data only had information on students who received federal student aid. In the worst instances, this meant results represented a minority of the enrollment at some colleges, excluding thousands of students. This led to the possibility of significantly underrepresenting completion rates or the extent to which institutions provide economic return to their students.

Under current federal law, however, the U.S. Department of Education (ED) has no way to fix these problems. This is due to a provision added to the Higher Education Act in 2008 that prohibits ED from creating a new database of information that collects individual-level data on all students enrolled in higher education—referred to hereafter as a student-level system. More commonly known as the "unit record ban," this provision only allows ED to operate databases that are necessary for the operation of the federal student aid programs and that existed prior to 2008.[9]

The unit record ban has far-reaching consequences that undermine the ability of policymakers, institutions, students, families, and the public to answer important questions about the effectiveness, efficiency, and impact of America's postsecondary education systems. For example, policymakers cannot judge the collective impact of our nation's colleges and universities in terms of meeting attainment needs or societal contributions. They also cannot answer crucial questions about the roughly $130 billion spent by ED on college assistance program each year, including federal student aid and other programs. This includes questions such as, "What is the federal government getting for its investment in terms of outcomes?" or "How might aid programs be better struc-

tured to ensure student mobility and close persistent gaps along racial or socioeconomic lines?"

Institutions also suffer from the ban on a student-level system. At a time when colleges are under substantial pressure to demonstrate their value, they struggle to capture and measure key outcomes on a national scale that would make a more compelling case for increased funding. They also have difficulty benchmarking outcomes to understand what should be reasonable expectations—resulting in outlandish initiatives from state leadership, such as Florida's goal of having community colleges graduate 100 percent of full-time students.[10] The extent to which benchmarking is a problem for institutions varies somewhat by state. Those with the most robust data collection apparatuses like in Virginia, Texas, or Florida may be able to answer many of these questions. But even the best state data systems still have three main drawbacks:

- ▶ They cannot track results for students who move across borders.
- ▶ They have limited understanding of outcomes associated with student loan debt.
- ▶ They have to rely on imperfect earnings information that may not capture all the positive results due to limitations on collecting data on federal employees, the self-employed, graduates who move out of state, and others.

Such limitations may be hindering efforts to make the strong case for the desperately needed state reinvestment to slow the inexorable rise in tuition and debt and ensure America's public higher education system can serve as an engine of growth and opportunity.

Arguably, the ban on a student-level data system hurts no group more than students seeking to enroll in college and their families. Without more comprehensive data, students are unable to answer basic questions such as, "Which schools do the best job serving students like me?" or "What is a reasonable amount of money to spend on a given institution or program?" This second question is particularly problematic since the ever-rising cost of higher education makes going to college an increasingly high-stakes financial decision. Without a proper assessment of the likelihood of finishing and other key outcomes like repaying loans and post-school earnings, students may be spending thousands of dollars for educational options that are not a good value for them.

To put all these challenges another way, ED hands out roughly $130 billion in federal aid each year, yet thanks to the unit record ban, everyone involved in the system—policymakers, institutions, students, families, and taxpayers—knows very little about what it is getting for its money.[11]

A student-level system at the federal level is the strongest solution for addressing the problems outlined above. It would create a powerful tool for policymakers to engage in smart and evidence-based decision-making. A student-level system would bolster institutions' abilities to make the case for their impacts at the state and federal levels and empower students and families to make improved choices that have lifelong implications for their economic wellbeing.

Establishing a federal student-level system would not require massive amounts of new information from colleges. For the 70 percent of students who receive federal financial aid, the system mostly exists. The remaining 30 percent, meanwhile, would be captured by having most colleges submit information to the federal government that they are already sending to a private third party. Institutions would then have to supplement these data with just a few new indicators. With these reporting changes, colleges would be able to generate and receive a wealth of information that would be extremely burdensome to receive through any other method.

Given all these potential benefits, there is increasing interest among policymakers, college trade associations, and the public for Congress to revisit the unit record ban. While making this happen requires political and strategic considerations, it raises policy questions as well: What functions must a student-level system perform? Who would manage it? How would students' privacy be protected? How much would it cost? What would it mean in terms of the reporting burden placed on colleges and universities?

This report tries to address those key policy questions. It looks at what the federal government already collects and what this might mean for a student-level system. It also discusses the benefits and drawbacks of several design options, with a strong focus on practicality. Ultimately, this report makes the case that a federal unit record system is a feasible option for improving American postsecondary data. In addition, the long-term benefits of such a system in terms of analytical capabilities, transparency, and a reduction in burden for institutions far outweigh the startup and maintenance costs.

The need for better data is particularly acute because the stakes involved have never been higher. Approximately 7.5 million Americans are currently in default on a federal student loan; millions more are delinquent and on the path to ending up there.[12] The cost of higher education and the increasing need to borrow raise the risk of attending college, making it even more important that consumers make good choices about where to go and what programs to pursue. In addition, in a world of limited federal resources, it is imperative that federal financial aid programs succeed in helping the largest number of students possible access and complete postsecondary education.

With Congress actively discussing the next reauthorization of the Higher Education Act, this is the time for revisiting the current federal policy landscape. In doing so, Congress should end the federal unit student record ban.

The Department of Education's Current Data Collections and What They Mean for a Student-Level System

Though the federal government cannot operate a student-level database that includes information about everyone enrolled in college, the Office of Federal Student Aid (FSA) already has one for all federal student aid recipients. Known as the National Student Loan Data System (NSLDS), this database contains 25.9 billion current and historical records on over 81 million students holding 363 million loans across over 19,400 schools (including branch campus locations and institutions that have closed or changed names).[13] It keeps information on student aid recipients until 15 years after they have paid off their accounts.[14]

NSLDS contains five main types of data. First is basic demographic information about a student—name, social security number, address, gender. Second, NSLDS includes records of a student's financial aid history. This includes the type and amount of aid a student received (e.g., a Pell Grant, Subsidized Stafford Loan, Graduate PLUS loan); when they received it; and the college they attended when they got that aid. Third, NSLDS contains information on when a student left school, including a flag that indicates when and whether a student graduated or withdrew. More recently, NSLDS also started requiring colleges to report what type of credential and program a federally aided student is seeking as a way of complying with requirements that limit the amount of time a student can receive Subsidized Stafford Loans.[15] Fourth, NSLDS includes a great deal of information related to student loan repayment. This includes indicators such as the loan holder or servicer, the loan's status, and balance. For Direct Loans it also includes information on the payment plan. Finally, NSLDS includes information from FSA's other databases. This includes information derived from the student's Free Application for Federal Student Aid (FAFSA), such as their expected family contribution, their income, and their dependency status. For dependent students, NSLDS also has data on the income, educational attainment, and demographic information of their parents at the time the student applied.

FSA uses NSLDS for operational purposes, particularly in terms of managing repayment and ensuring the integrity of the aid programs. For instance, federal law requires that borrowers must start repaying their loans six months after they graduate or leave college; knowing the date a student graduated or withdrew is extremely important. Similarly, the level of aid received is important for ensuring that students received the proper amount of funds. NSLDS also generates certain accountability metrics required by Congress, such as the percentage of students who default on their loans within three years of entering repayment. Finally, FSA also uses NSLDS to link to other parts of the postsecondary data infrastructure. For example, data from NSLDS were linked to the U.S. Department of the Treasury's administrative earnings records to generate estimates of the income of federally aided students on the College Scorecard.[16]

Outside the federal government, NSLDS's most notable connection is to the National Student Clearinghouse (NSC). This is a private third-party database to which colleges regularly report information on their enrolled students, including degree attainment and other indicators. . Many colleges rely on NSC to prepare the enrollment reports they must send to NSLDS. To do this, they will submit to NSC a roster of information on all their students semi-regularly (about once a month or roughly four times per term).[17] NSC then matches that data to rosters of students sent to it by NSLDS, fulfilling the institution's requirement that it notify ED about the enrollment status of its federally aided students.

While NSLDS is the best-known database run by FSA, that office administers two other systems that merit mentioning. (For a lengthier consideration, please see Matthew Soldner and Colleen Campbell's paper, *Using—And Improving—FSA Data Systems to Support Policy Analysis*.) First is the Central Processing System (CPS). This database holds the information students submit when applying for federal aid, including data such as a student's dependency status, income, parental information (if they are dependent), and other similar characteristics.[18] CPS also stores the calculations run on a student's aid application, such as their expected family contribution. Much of the data generated and held by CPS are later added into NSLDS. Second is the Common Origination and Disbursement system (COD). FSA uses this database to send financial aid to schools.[19] It stores detailed information about the aid awarded to a student, as well as demographic information, their enrollment dates, attendance intensity (i.e., full-time or part-time enrollment), and level in college. With NLSDS acting as the long-term storage of financial aid information, COD serves as the database for short-term and immediate change, such as adjusting the amount of aid sent to students or requesting additional funds for an individual.

Separate from the FSA databases, ED's National Center for Education Statistics (NCES) administers a separate database of college information known as the Integrated Postsecondary Education Data System (IPEDS). Unlike FSA data, which ED collects at the student level, almost all IPEDS data are at the institutional level, with a few indicators such as the number of graduates and some pricing data also reported by program. Colleges are required to report information into IPEDS as a condition of participating in the federal student aid programs. The data they report cover a wide range of topics in five main areas. First, there is basic information about the college, its tax status, location, types of programs offered, etc. The second topic area is pricing information, including the listed (or sticker price) and the net price—that is, how much full-time freshmen attending for the first time pay after subtracting grant aid—and the amounts of financial aid received. The third area is outcomes data, such as graduation and retention rates, as well as the number and type of degrees and certificates awarded by program. Fourth is enrollment information, including the number of students enrolled in the fall and over the entire academic year. Finally, IPEDS data includes information on an institution's staffing and financial indicators. This includes balance sheet data, such as revenues, expenses, and endowment assets, as well as counts of faculty and other staff.

Because IPEDS is an institutional-level data collection, it obtains information in a different manner from FSA data. Data in NSLDS come from several sources: FSA, the institution, and student loan servicers, among others. This allows multiple parties to feed in information simultaneously. IPEDS solely uses institutional reporting, which in some cases a state may also complete on behalf of multiple colleges or systems. IPEDS also relies on quarterly data collections, while FSA data update constantly.

Beyond these continuous (FSA) or annual (IPEDS) data collections, the Department of Education also administers student-level sample surveys through NCES every few years. These include:

▶ The National Postsecondary Student Aid Study (NPSAS), a quadrennial survey on the status of financial aid across all students in college, ranging from incoming first-year students all the way up through graduate education, last administered in 2011-12[20]
▶ The Beginning Postsecondary Students (BPS) Longitudinal Study, which tracks completion, transfer, credit accumulation, course taking, and other issues for a sample of students one, three, and six years after first enrolling (the latest BPS survey came out in 2016, which measured outcomes after three years for students who first enrolled in 2011-12)[21]

▶ The Baccalaureate and Beyond (B&B) Longitudinal Study, which tracks bachelor's degree graduates after finishing to capture information on jobs held, further studies, and other topics. (the most recent B&B data release came in 2014, which tracked data for students who graduated in 2007-08 after four years)[22]

The idea behind these surveys is to generate a nationally representative picture of higher education by conducting extensive samples with tens of thousands of students. These samples abide by NCES's rigorous methodological standards. Because they are samples, however, the surveys cannot generate individual results by college, and rarely can break results down by state. They do, however, represent a level of comprehensiveness that is not currently replicated in any other federal collection. They also have robust public data analysis tools that allow users to create customized queries.

Why Existing Federal Collections Are Insufficient

The combination of NSLDS, other FSA databases, IPEDS, and NCES sample surveys provides a wealth of information about American higher education. These data systems also have significant drawbacks. To be clear, these problems are less about the systems themselves and more the fact that the purposes they serve and today's data needs no longer align. For NSLDS and other FSA data systems, the issue relates to what data are collected and who is included in the database. For IPEDS, the problems result from the burdensome costs and staff time needed for institutions to report data and the limitations of an institutionally based reporting method. For the NCES sample surveys, their lack of state and institutional-level results and irregular collection make them insufficient to meet policymaker and student needs.

TABLE 1. FEDERAL STUDENT AID (FSA) AND NATIONAL CENTER FOR EDUCATION STATISTICS (NCES) DATA COLLECTIONS

Location within the U.S. Department of Education	Data System	Collection Level	Types of students included	Indicator Types
FSA	National Student Loan Data System	Student	Only federal aid recipients	Student demographics Federal aid amounts and history Enrollment history Loan status
FSA	Central Processing System	Student	Federal financial aid applicants	Student demographics Student income, assets, educational attainment If applicable, parental income assets, and educational attainment Expected family contribution
FSA	Common Origination and Disbursement system	Student	Federal financial aid recipients	Student demographics Aid amounts, dates, and types Program and institutional information
NCES	Integrated Postsecondary Education Data System	Institutional	All students (only first-time, full-time for most outcomes)	Institutional characteristics Graduation rates Institutional pricing Enrollment Admissions Institutional finance and staffing
NCES	National Postsecondary Student Aid Study	Student	Nationally representative sample of all undergraduate and graduate students enrolled in a given year	Student demographics Institutional characteristics Amounts of financial aid from all sources Enrollment status
NCES	Beginning Postsecondary Students	Student	Nationally representative sample of all undergraduate students starting college in a given year, tracked for up to six years	Student demographics Institutional characteristics Attainment, withdrawal, transfer, and other outcomes
NCES	Baccalaureate and Beyond (B&B) Longitudinal Study	Student	Nationally representative sample of all students graduating with a bachelor's degree in a given year, tracked for several years after finishing	Employment outcomes, including earnings, retirement savings, and benefits Student loan status in repayment Graduate school attendance

The challenges of NSLDS and other FSA databases are the most evident. For one, the systems only include data on students who received (or in the case of CPS, applied for) federal aid. This means that about 30 percent of all graduating students are not in the database.[23] While these gaps are particularly acute at public 2-year institutions, where 38 percent of students receive no federal aid, they are a major problem in every sector of higher education. Many individual colleges have even higher percentages of students not receiving federal aid.[24]

In addition to not covering large groups of students, NSLDS and other FSA data also do not have certain key indicators that might be of use. For instance, they do not collect information on a student's race, whether a student was college ready, or how many credits a student completed. FSA also does not know anything about non-federal financial aid the student received, so the system cannot generate estimates of measures like net price or unmet need.

While IPEDS includes all students, not just those receiving aid, it has other challenges. First and foremost, because institutions report aggregate numbers, adding new indicators means the college must report a new set of data and a new cohort each time. For instance, IPEDS currently requires graduation rates by race and ethnicity. To generate these figures, the institution must report separate cohorts and numbers of graduates for each category. It cannot simply submit a roster of students and then query it to filter results based upon categories needed. This means adding any disaggregated figures will require colleges to run new calculations, increasing time and cost burdens.

IPEDS's second major challenge is that federal law requires that many of the potentially useful indicators use a cohort that is not tailored to today's students. Federal law requires colleges to report a specific graduation rate to IPEDS: someone who enrolled as a full-time freshman at a given college and had never enrolled elsewhere. Limiting completion rates to these so-called "first-time, full-time" students excludes several sizeable groups of students. First are those who transfer into a college—over 37 percent of students.[25] Second are students who enroll part time—about 38 percent of people in college, including nearly 60 percent of those at two-year public institutions.[26] In addition, many colleges may have large numbers of students who do not start in the fall.[27]

To its credit, NCES is fixing some of these drawbacks. It recently added new measures to IPEDS requiring colleges to provide data on part-time students enrolling for the first time, as well as students who transfer in and attend either full or part time.[28] In 2016, NCES announced that it would further expand the collection to include completion rates for students receiving Pell Grants.[29] These are important indica-

tors, but it is not feasible to disaggregate them in all the ways needed because doing so would add an unreasonable amount of reporting. For example, if NCES wanted to start collecting information on the graduation rates of part-time students by race and gender, it would be asking schools to report approximately 30 new graduation rates—and that is just for one new cohort type.[30]

The NCES sample surveys simply are not designed to fulfill the role of a student-level data system. The sample surveys cannot generate results for specific institutions, rendering them less useful for helping students, policymakers, and college leaders make decisions. The sample surveys are also not timely. Conducting a rigorous and high-quality survey takes a great deal of time and costs a lot to run. Because of this, institutions often only collect the most common survey every four years, and others may go much longer between administrations. This means data may not appear fast enough to understand rapid changes in higher education.

In the short-term, there are potential fixes to both FSA or IPEDS that would improve these systems and address some of the challenges noted above (see Matthew Soldner and Colleen Campbell's paper, *Using—And Improving—Federal Student Aid Data Systems to Support Policy Analysis* for more on FSA improvements and Jamey Rorison and Mamie Voight's paper, *Putting the "Integrated" Back Into IPEDS: Improving the Integrated Postsecondary Education Data System to Meet Contemporary Data Needs* for more on IPEDS). While these improvements are valuable, they cannot undo major structural hindrances to these systems. For example, improvements to FSA data cannot eliminate the statutory ban on creating a student-level data system that prevents the compilation of information on everyone enrolled in college.[31] Similarly, new IPEDS indicators cannot make a system designed as an institutional information collection fulfill the needs of a student-level data system.

Why State and Institutional Data Systems Are Insufficient

The federal government is not the only actor within the higher education system that faces substantial gaps in its data collections that hinder effectiveness. Institutions and states also face challenges, albeit of different sorts. This section lays out some of those problems in general terms. For a more specific discussion, including some indication of how these problems may vary, see Randy Swing's paper, *Institutional Research Capacity: Foundations of Federal Data Quality* for a discussion of institutional data and John Armstrong and Katie Zaback's paper, *Assessing and Improving State Postsecondary Data Systems* for a consideration of state data systems.

Institutional data has the benefit of being more nuanced and detailed than any other higher education data collection. Col-

leges, for example, could choose to collect data on college readiness, courses taken, attendance records, grades, or a host of other topics that go well beyond what states and the federal government know or need to know about a given student. These detailed data collections give institutions of higher education greater insight into nuanced discussions, such as measuring the effectiveness of classroom interventions or student supports, as well as understanding student learning outcomes (to the extent that they measure them). Colleges are also the only ones that will know about any institutional aid they awarded to a student.

Institutional data comes up short, however, when students start moving around. While a college knows a great deal about someone enrolled on their campus, it is unlikely to know much about that person when they leave. For example, when someone transfers out of a college, the school may be able to find out if they enrolled elsewhere, but it might not know what they are studying or how many credits transferred. Similarly, when someone graduates, the institution might get reports on whether that person is paying their loans, but it will have a more limited picture of that student's long-term debt results or workforce outcomes, especially if that individual moved to another state.

State data systems are capable of knowing a lot more about when students move around than an institution's data system. By collecting information from multiple colleges, a state can better track who is enrolling where and understand what happens when someone transfers from a two-year to a four-year college. Many states also have the ability to generate some workforce information by linking to their unemployment insurance system. See Rachel Zinn's paper, *Classroom to Career: Leveraging Employment Data to Measure Labor Market Outcomes* for more on workforce outcomes and higher education data. In many places, state data systems also link back to K-12 schools, making it possible to assess factors such as how well graduates of various high schools do in college.

States, however, also struggle with tracking student mobility, particularly across borders or often into private colleges. Only a few state data systems include information from private non-profit colleges and almost none cover private for-profits. Even a state system that covers every single college in its borders is going to run into trouble measuring outcomes when someone moves across state lines. In addition, state data systems do not contain long-term information on how students handle federal loans or other things related to the federal aid programs.

Fortunately, there is a solution that can fix the challenges of insufficient federal, state, and institutional data: a student-level data system.

A Student-Level Data System

A federal student-level data system is a database maintained by the federal government that contains information at the individual level for everyone enrolled in college. Only by collecting and maintaining these data at the student level, and not by institution or some other categorization, does it become possible to create a system capable of addressing the various limitations and gaps that hamper existing federal, state, and institutional data systems. Notably, it is also the only way to accomplish this goal through a manner as close to automated as possible, minimizing the added burden on everyone involved.

The Benefits of a Student-Level Data System

The current federal higher education data landscape is like a puzzle— various actors each have some the pieces but no way to put them all together to form a finished product. This creates information gaps that hinder analysis by states, institutions, and the federal government, and prevents students, the public, and policymakers from getting the transparency they need. A student-level data system creates a structure for bringing the puzzle pieces together by allowing for better data sharing.

Here is an example of the type of information gap that better data sharing could fill: Institutions currently know very little about workforce outcomes or long-term loan performance of their students. The federal government knows the latter and is capable of generating information on the former but needs to know which students go into which cohorts—for example, who attended and graduated from a given college in a specific year. Fortunately, that is something the school knows very well. Under a better system, these two stakeholders could essentially trade data—the school tells the federal government who graduated and the federal government uses that to return more robust workforce information. Both parties benefit: the federal government learns more about completion and the institution gains knowledge about earnings.

The benefits of a student-level system, however, are about much more than the flow of information. The potential uses of those data are what makes the student-level system such a powerful option. Below are more details on how the system would benefit everyone involved in the postsecondary data system.

Institutions. For institutions, a student-level data system would be a powerful tool that would let them better document success, gain new information on outcomes, assist in efforts to advocate for additional funding, and potentially help with recruitment. On the first issue, many institutions, especially community colleges, struggle to capture all of the positive outcomes they produce through current data. This is particularly an issue around graduation rates, as described

above, since traditional measures do not sufficiently give colleges credit for students who transfer out and complete or for those who transfer in and finish. Having a student-level data system would ensure that institutions would get credit for all their graduates from the federal perspective.

A student-level data system would also grant institutions access to outcomes information that they currently have significant difficulty obtaining. By linking data they send to other federal sources, institutions would be able to learn more about the workforce outcomes for their students. They could more easily answer questions such as, "Do students go on to additional education and training, and if so, what types?" Colleges could also learn much more about how their students are handling federal loan debt.

With better information, institutions will also gain more tools that can help them make the case for continued or increased funding at the state or institutional level. In addition to capturing all of the graduates they produce, institutions could use the output from a student-level data system to document the economic gains associated with their programs, strengthening the case for postsecondary investments as a tool for growth.

Better data can also help with program improvement, recruitment, and retention efforts. By knowing accurate and complete earnings results for students by program, a college can gauge which offerings are worth sustaining and growing and which ones might need adjustments to increase quality. They can similarly use those results to help recruit students by showing the value of their programs. If colleges can disaggregate results by those who did or did not complete, they can make the case to students who are thinking of dropping out and taking a job that staying in school will provide better long-term benefits.

Finally, a student-level data system could help institutions more easily fulfill existing requirements related to data reporting for IPEDS. This issue will be discussed more in a later section, but combined, colleges could save over a half a million hours a year in IPEDS reporting by instead sending data to a student-level system (see estimates on **page 18**).[32] Reporting data through a student-level system would also make it easier for institutions to adjust to new metrics. Currently, every change in IPEDS data definitions requires colleges to redo their reporting structures. By contrast, a student-level data system can easily change indicator calculations without any additional burden on institutions, since the results would just be sliced in a different manner. To understand how this is better, imagine that the higher education community decided to redefine the graduation rate to include all full- and part-time students. To make this work in IPEDS, schools would have to construct and report entirely new cohorts. Doing this under a student-level system would just mean programming the database to run a different query.

Students and families. A student-level system would grant students and families access to powerful new sources of data that could help them make better choices about college. For instance, students would be able to generate more personalized information by program or by institution related to completion, repayment, and earnings outcomes for people similar to them. This would allow a 26-year-old, low-income Latino student to understand the exact outcomes for someone of a similar background who attended a particular institution rather than getting the current statistics that include people with little in common with them demographically. Knowing this information would close some of the significant information gaps that exist now, helping families make better judgments about proper amounts to pay or borrow for college. It would also exert greater market pressure on colleges by strengthening consumers' ability to vote with their feet and their dollars.

Furthermore, this data system would still be of great help to students who do not have the luxury of choosing from among multiple colleges. An increasing body of research is pointing out that many students, particularly lower-income ones, are geographically constrained and may have few college options to consider. For these individuals, data on how different institutions perform may not be helpful. Nevertheless, these students still have to choose a major within the school, and in that situation, the student-level data could help them make more informed choices about majors and programs.

States. States, too, would benefit from a student-level system. As discussed earlier, even the best state systems are typically limited in three main ways: (1) coverage beyond their borders, (2) data on financial aid performance, and (3) coverage of all private colleges. A student-level data system could correct for each of these challenges. National reporting on all students would allow states to understand what happens when students migrate outward, helping them get a fuller picture of where their students go and how successful institutions are. In terms of financial aid, states could learn two key things. First, they could better evaluate the effectiveness of their own aid programs, using the federal system to see the interaction between state aid and borrowing, as well as earnings and completion. Second, they could better understand the level of indebtedness of students in their state, information that may be useful in considering the need to tackle college prices. Third, getting data on all private colleges will help states get a complete picture of everyone enrolled in an institution of higher education in their borders.

Federal government. Finally, the student-level data system will also help the federal government better assess the state of American higher education. Much like states and institutions, it would have a more complete picture of outcomes, making it easier to identify pressing national gaps. It can also use these data to better steward taxpayer dollars spent on

federal student aid programs. The federal government would be able to contextualize the results for students receiving federal aid versus those who do not, allowing it to understand if federal programs achieve desired results and to identify potential policy changes. A federal student-level data system would also benefit the government from a societal standpoint, allowing it to track higher education's effectiveness in promoting social mobility and economic growth.

To put all the benefits above another way, a federal student-level data system would bring down the barriers of data opacity that currently create problems throughout the system. It would level the power imbalance between institutions and students, which leaves the latter in the dark about what they can expect from their school. It would also place institutions on stronger footing with their states, allowing them to better capture their results and demonstrate their worth. The federal government, too, would find itself no longer playing the role of the rich uncle whose checks are cashed without a greater understanding of what the money actually buys.

Why We Do Not Have a Student-Level Data System Now?

The idea of a student-level data system is not without controversy. In the past, there has been a substantial split among the organizations representing colleges about this idea. While some, such as representatives of public colleges and financial aid administrators, have been supportive, others have opposed the proposed system.[33] The dynamics around this issue are also constantly evolving, as states, institutions, the federal government, families, students, and the public increasingly recognize that existing postsecondary data often falls short of today's information needs.

Some of the objections to a student-level data system represent a fundamental difference of opinion about what data the federal government should be able to know about students in college. There are some institutions receiving federal funds that do not believe that this support should give the federal government the authority to establish further requirements for disclosure about outcomes or other key data. How they come to this belief is not clear. In some cases, this may be due to privacy considerations; in others, it could be about not wanting to shoulder reporting burden or simply wanting a limited federal role. Even individuals not opposed to the student-level ban for any of the reasons outlined above may still have concerns about data security.

While it may be difficult to find common ground on some of the concerns above, stakeholders could resolve other disagreements simply by fostering greater clarity around the student-level data system. Without a specific plan for what the system would need to do, it becomes easy to generate overstated hypotheticals that include indicators the federal government would not want, such as detailed records on a student's health or disciplinary history. It also makes it difficult to judge the actual burden for institutions relative to the potentially significant benefits for them as well as for students, families, policymakers, and states.

The next section of this report adds clarity to the student-level data discussion by breaking down some of these key questions to scope out what features and data elements such a system would need to include. This also entails considerations of interoperability, privacy, and data access.

The Necessary Functions of a Student-Level System

Determining necessary functionality is the first step in scoping out a student-level data system. Knowing what the system needs to do will guide the indicators collected, the student disaggregation needed, the frequency of data collection, and the interplay between a new student-level system and existing data reporting requirements for NSLDS and IPEDS.

A student-level data system must have six main functions (see **Table 2** below for a simplified breakdown of these functions). First, it must be able to construct a complete picture of everyone in college. This is the prerequisite for generating

TABLE 2. THE SIX NECESSARY FUNCTIONS OF A STUDENT-LEVEL DATA SYSTEM

Function	New reporting required?	Who reports?
Construct a complete picture of everyone in college	Data on non-federally aided students, credit accumulation, level in college	Institutions
Collect information on non-federal student financial aid and price	Data on non-federal grants and loans for all students and prices charged	Institutions and possibly states
Disaggregate data by major student subgroups	Race, college-ready status	Institutions
Link to National Student Loan Data System and other federal data (e.g. earnings sources)	None	Department of Education, Department of the Treasury, Social Services Administration, Department of Veterans Affairs, Department of Defense
Report data back to institutions and states	None	Department of Education
Publicly release data, including through a customizable interface	None	Department of Education

the nuanced and full pictures of student completion, transfer, workforce outcomes, and financial aid results that make the system worthwhile. Second, it must be able to produce more accurate measures of financial aid to help stakeholders better understand how federal investments are leveraged and what students actually pay. Third, it must be able to disaggregate data by important categories, including demographics and programs within an institution. Fourth, it must link to data other federal sources hold, including post-college earnings. Fifth, it must be able to provide data back to institutions and states for easy use and analysis, particularly to disclose data on topics such as federal student loan performance that the schools cannot otherwise obtain. Finally, it must make data available to the public. This includes the release of analyses as well as an interface capable of allowing for aggregated analysis of student data akin to what users can currently do with the NCES sample surveys.

The following sections discuss the rationale for each of these functions and their implications from a technical and data collection perspective. The first three functions represent the minimum data points needed in the system to make it capable of producing useful analyses. The second three speak to the necessary linking and reporting functionalities the system needs in order to be useful to institutions, states, consumers, and the federal government.

Function 1: Construct a Complete Picture of Everyone in College

A student-level data system must contain a comprehensive picture of who is in college and what happens to them. This kind of foundation is necessary for properly constructing cohorts of students for measuring outcomes. Complete coverage ensures that linking student-level data to other systems, such as those that generate data on earnings or financial aid, produces results that are accurate and capture as many people as possible.

Building the foundation from which the rest of the student-level data system can operate requires collecting four types of data: (1) who enrolled, (2) when they enrolled, (3) where they enrolled, and (4) what happened to them. The first two data items are crucial for properly defining cohorts and tracking students for the right amount of time, the third dictates which schools should include that student in their metrics, and the last indicates which students to use when breaking down results by completers and non-completers.

The types of data described above have to come from either individual colleges or state data systems because the federal government currently only has limited information on enrollment and completion. For federally aided students, the government knows if a student graduated based on a flag in NSLDS completed by an institution's financial aid office. For

some cohorts of students the government knows the graduation rate based on IPEDS reports. These data, however, are limited and incomplete. For instance, only 70 percent of students receive some type of federal financial aid, including nearly 40 percent of students at community colleges.[34] Some estimates place the number of students covered by the IPEDS graduation rate at below 50 percent.[35] The current graduation rate formula treats students who do not fall into either collection as if they never existed, and IPEDS in particular struggles with handling mobile students who attend multiple colleges and disaggregates these data in few ways beyond race and gender.

Colleges are already reporting the above four types of data on federally aided students. For these individuals, the colleges have to note their enrollment dates and their level of attendance intensity through regular reporting to NSLDS. In addition, a statutory change in 2012 required colleges to start providing information on the program type and degree level. ED also recently started stressing that colleges indicate if someone graduates or withdraws. This includes the addition of a flag in July 2012 that identifies Pell Grant recipients who graduate.[36]

Many colleges also already do this type of reporting on their students who do not receive federal financial aid to the National Student Clearinghouse (NSC). Currently about 4,500 colleges fulfill their NSLDS reporting by sending data to NSC.[37] In those instances, they typically report data on all their students, not just those receiving aid. Importantly, the data they already report include optional variables for race/ethnicity as well a remedial education flag—ways of disaggregating data that should be added to a student-level system.[38] NSC matches a college's data to a list of students receiving federal financial aid from NSLDS. If the student-level data system links to NSC, colleges may be able to expand their reporting to the federal government to cover all students with minimal changes to their current practices.

While the information outlined above will make it possible to generate much better completion measures, two additional types of data would significantly improve results. First is a picture of students' progress through postsecondary education as measured in credits. This includes two measures: the total number of credits attempted in that term and the total number of credits earned in that term. Combined, these two measures can be used to derive information on total credit accumulation. This information is crucial for any analysis of efficiency, such as measuring if students earn excess credits on the path to completing a degree. These data could also potentially serve as an imperfect proxy for whether a student should be counted in completion rates because they are seeking to complete a credential at an institution. Credit accumulation also matters for judging whether students who

transfer and complete end up taking on more credits than those who do not. Second is the student's level in college. Students currently report this measure on the FAFSA, but it may not be accurate because they may not understand that they have not completed enough credits in a year to be considered a second-year student or some similar issue. Institutions report this information for loan purposes into the Common Origination and Disbursement (COD) system but would need to do so for all students. Knowing a student's level in college matters for defining proper cohorts and judging transfer success (so a college can see what level a student ends up in at another institution).

Function 2: Measure Student Financial Aid and Net Price
Collecting information on non-federal financial aid is the second base component for a student-level data system. This matters for several reasons. First, knowing the complete picture of federal, state, and institutional grant aid helps the federal government understand whether the students it helps still face gaps in what they have to pay for college, and if so, how much. Including non-federal loans also helps the federal government understand the interaction between its debt products and other assistance students receive. States, meanwhile, would benefit from these data because it would allow them to measure student earnings, completion, or other outcomes disaggregated by whether or not a student received support from one of their grant programs. Finally, these data would provide students with more personalized estimates of their net price—that is, what they will likely pay out of pocket after deducting all grant aid. This latter piece of data is very important to students. Two of the three most important factors identified by current and prospective students when picking a college related to cost are the availability of financial aid and the price charged.[39]

Institutions already report aggregated information on non-federal financial aid in the IPEDS Student Financial Aid component. However, the data are limited to aid received by a cohort of entering full-time, first-time degree/certificate-seeking undergraduate students. As a result they fail to capture information on large numbers of students.

Collecting details about non-federal aid and pricing would make it possible to present students and consumers with a far more accurate estimate of net price than what exists today. While colleges are currently required to report a net price into IPEDS, that calculation is incomplete at best. For one, it only provides estimates for students who are attending full time and have not enrolled elsewhere. This brings up the same issue present in the graduation rate data: it excludes large numbers of students, especially those going part time or who transfer in. In addition, the net price only presents information for students in their first year. Most students attend college for several years; understanding whether the

amount they pay in the first year is likely to change would also substantially assist families in making choices. None of these problems would exist in a student-level system. Having financial aid information at the individual level would make it possible to generate net prices for students who are part-time or transferred in, among other characteristics. Collecting non-federal financial aid information at this level would also make it possible to generate a longitudinal net price measure that reflects the total amount paid over an entire time in education, not just the first year.

Function 3: Disaggregate Data by Important Subgroups
The final set of data elements required for a student-level system includes indicators to identify whether a student is a member of subgroups that are important to consumers and policymakers. Collecting this information will make it possible to break down information by race/ethnicity, income, college readiness, gender, age, military or veteran status, and first-generation status. This is necessary for ensuring that vulnerable populations get sufficient attention from the institution. The advantage of a student-level system is that as long as the institution reports whether or not a student has a given characteristic—such as not being college ready, being in their first year, etc.—it is possible to generate multiple cohorts without reporting them all separately. In other words, it is possible to calculate a completion rate of Pell-eligible students who are not college ready versus all Pell-eligible students without needing to report two separate cohorts of students.

Some of the necessary types of disaggregation are standard demographic information that colleges would expect to report in any data submission. This includes characteristics such as gender, race/ethnicity, and age. The enrollment and completion data already contain other disaggregates. This includes indicators such as the type of credential sought, level in college, the student's program, and attendance intensity. Linking to NSLDS can also provide disaggregates such as whether a student received a Pell Grant, Subsidized Stafford Loan, or other form of federal financial aid.

College readiness status is the most important disaggregate that is not available from other sources. Students who are not ready for college-level work frequently take remedial or developmental courses. Remediation is a large problem: one-third of students end up in these non-credit-bearing courses, including 41 percent at public two-year colleges.[40] These courses do not carry college credit and an estimated 40 percent of students never pass them.[41] Knowing the extent to which students are not ready for college will make it possible to identify where certain programs may or may not be working. A simple flag that asks colleges to report whether or not they deemed to a student to be college ready, which could include something as simple as whether a student was ever

required to or did take a remedial or developmental course, would fulfill this data need.

Admittedly, the data collected from a college readiness flag will be of lesser quality than other elements. That is because there are no standard definitions of what readiness means and the policies institutions use for making that determination varies. Some places, such as Florida, have made remediation optional for students, resulting in basically everyone appearing to be college ready.[42] As a result, a student who is college ready in one institution may very well not be in that situation at another, or vice versa. Additional work by NCES could also address this problem by convening a panel to define this college readiness flag—something already called for in 2011.[43]

One other type of disaggregation merits special discussion from a data collection standpoint: financial and parental information upon entering college. These data matter because they provide important information about a student's economic circumstances, which helps judge how colleges do in serving low-income students and gauge if higher education is successfully serving as an economic mobility tool. These data, however, are currently limited and only available for students who fill out the FAFSA. As a result, these data are unavailable for a significant population of students. According to the 2011-12 National Postsecondary Student Aid Study, 20 percent of all students never applied for financial aid.[44] This includes 30 percent of students at public two-year colleges.[45] The system would need to either find a way to obtain these data from some other sources—such as a state or an institution—or rely on a proxy measure like a student's zip code.

Function 4: Link to NSLDS and Other Federal Systems
The first three functions describe what needs to go into a student-level data system to provide sufficient information to measure key outcomes and sort students into important groups for judging success. To make that additional data reporting by institutions and states worthwhile, the system must also be able to query and receive data from other federal systems. These federal linkages also allow as many indicators as possible to come from non-institutional sources, reducing the amount of information colleges must report. In all cases, these types of connections must be done in ways that follow applicable privacy rules and closely guard individuals' information. In making such linkages, the system should also ensure that it only receives indicators that are absolutely necessary and does not store information that serves no purpose.

A student-level data system must be either built in or be capable of linking to NSLDS. This is necessary for both operational and analytical reasons. From an operational standpoint, linking to NSLDS provides a significant opportunity to minimize the additional burden on institutions. FSA requires colleges

participating in the federal aid programs to submit updates at least every two months on the enrollment status and program choice of federally aided students.[46] The data contained in this process represent most of the indicators necessary for a student-level system.[47] Therefore, any reporting for this new system should be able to fulfill the existing NLSDS enrollment reporting requirements. This includes collecting the same indicators on the same reporting cycle. Doing so makes it possible for colleges to avoid reporting the same information twice.

The student-level system must also be able to connect to other major federal databases. In particular, it must be able to include three major types of data: other FSA data, earnings information, and veterans' benefits.

As discussed earlier, FSA maintains two other systems, namely the Central Processing System (CPS) which houses data on people who apply for federal aid and the Common Origination and Disbursement (COD), which allows schools to initiate requests for funding for financial aid disbursements. While NSLDS already houses the most important indicators in these databases, it would still be beneficial to link the database to these other systems in case they have other information that would be useful to help with data disaggregation—such as income information for people who apply for but do not receive federal aid.

In the past, the Department of Education has relied upon two different sources for obtaining earnings data. The first is the Social Security Administration (SSA). ED already has experience using SSA data for judging the earnings of graduates who received federal aid at career-training programs as part of the gainful employment regulation. Under that rule, ED sends SSA rosters of students who received federal financial aid grouped by the institution and career-training program within it they finished. Each list must have at least 30 student names on it. SSA sends back to ED a count of the students that did not match as well as mean and median earnings levels. However, SSA never provides ED with earnings data for individuals and has strict rules to protect privacy that have analytical implications. SSA requires a difference of at least 10 matched names before it will run data on a group of individuals for which it already generated earnings information in the same year.[48]

Consider the following example to understand how the policy on at least 10 different names affects earnings results. ED first sends SSA a list of 48 graduates from a cosmetology program at a given institution to get their earnings for the 2014 year. All those names match and SSA receives mean and median earnings data. ED wants to know the earnings information in the same year for the 40 students within that group who received Pell Grants. SSA will not process this information because the number of changes between the two ros-

ters—eight—is under ten. In this case, ED cannot obtain both the typical earnings of graduates and the earnings of only Pell recipients from that program. However, ED could submit this different roster of students to get earnings information for the 2013 year with no problem if it did not already request that year of data from SSA.

Administrative data from the Department of Treasury provide the second path to retrieving earnings information. This is the data source ultimately used by the College Scorecard.[49] While these data should be as reliable as SSA, they also have the same restrictions on generating results for individual students and requiring at least 10 changes to the roster of students before it will run a new query.

Regardless of whether the information comes from SSA or Treasury, earnings data held by the federal government are likely to be much more comprehensive than similar information obtained from other sources. Because states administer them, unemployment insurance databases do not cover someone who moves to another state. Nor do they include federal employees, individuals in the military, or those who are self-employed. Federal data do not have such shortcomings.[50] For example, SSA matched 99 percent of the student records sent by ED as part of a suggested College Scorecard analysis in 2013.[51] By contrast, research has estimated that only about 82 percent of workers are in jobs that would show up in unemployment insurance databases, including presumably large numbers of people who may be working in a state different from the one where they attended college.[52] While the exact coverage of unemployment insurance will vary by state, the number of missing individuals can be quite high. When the Western Interstate Commission for Higher Education created a four-state data exchange to improve coverage of outcomes measurement, it could only find data for about 60 percent of bachelor's degree graduates and 66 percent of associate degree graduates in their home states.[53]

The lower match rate does not mean state earnings data cannot serve a useful purpose. In particular, state unemployment insurance data include codes describing the industry and occupation of employees, and administrators of these systems may be more willing to report earnings information for individual people. This could be very useful for states or institutions to get a better sense of graduates' job placements and to identify individuals who may be struggling. From the federal perspective, however, it may be best to use national data.

It is also important to link the student-level database to information on current and former members of the armed services. Individuals who are contributing or have contributed to the country by serving in the armed forces should be able to

know whether the institutions where they can use their Post-9/11 G.I. Bill, Department of Defense (DoD) tuition assistance benefits, and other support will help them succeed. Including data from the VA and the DoD also makes it possible to generate completion rates for veterans without any additional institutional reporting.

Function 5: Report Data Back to Institutions and States
Simply collecting information for a student-level data system is not enough. If it just sits in a database, it will have limited value to institutions and states. Creating a data system that is actually used will also help with information quality, since institutions will be more likely to catch mistakes and pay more attention to the initial accuracy of reported data.

From an institutional standpoint, the student-level data system must have three main capabilities. First, it must be able to provide pre-designed reports on key issues that will help institutions manage accountability requirements for the federal student aid programs as well as improving their own outcomes. These reports could take several different forms, but the overall goal would be to send information back to institutions that they could not otherwise obtain. For instance, colleges could receive a loan repayment report, which combines NSLDS and student-level data to present characteristics of borrowers who are delinquent or in default on their loans. While that report could not contain individual earnings information, it could include aggregated income information for borrowers who are in default or distress.[54] This could identify patterns, such as non-completion, credit accumulation, or other factors that could target future outreach efforts. Similarly, reports on college graduates by program could help institutions make choices about what to offer. These reports may be of particular use for under resourced institutions that lack the capacity for their own internal data analysis.

Second, institutions must be able to judge transfer and completion patterns of former students. While many institutions may already get these data from NSC, the federal system must be able to supplement information on where an institution's former students enrolled with additional indicators. This includes indicators such as aggregated mean or median earnings data for all transfer students, detailed information about individuals who borrowed at that institution and elsewhere, and their performance on that debt.[55] This latter type of information should be of particular use to community colleges, which often must figure out how to help former students who have minimal to no debt from their institutions and who then go elsewhere and borrow a great deal more. Finally, because all institutions in the federal aid programs would use this student-level data system, it would have more universal coverage than NSC, in which colleges voluntarily choose to participate.

Finally, institutions must have the ability to generate their own personalized queries about their current and former students, such as results for students of color who borrowed. This is a major existing limitation of NSLDS. Colleges can only receive certain pre-determined reports. Those who want to learn more about their students must pull individual data.

States have similar needs to institutions, but with a slightly different perspective. Some may want data back on the performance of students who received state financial aid. This could include information on the completion status, mean or median earnings, student loan performance of state grant recipients, or perhaps data on the extent to which these individuals still face unmet need. In other cases, a state may want a greater understanding of how its overall systems of higher education perform. This could mean aggregating data for all public colleges of a similar type within a state (such as predominantly bachelor's degree granters), or also looking at data for individual public colleges. Finally, states may wish to query the federal data system to understand the outcomes for their high school graduates. This would help a state understand the pattern of where their former K-12 students go to college and help with the generation of feedback reports for high schools. In the states where the data systems are already robust enough to do this, such queries would be most useful for filling in the gaps to capture students who enroll out-of-state or attend private colleges.

Function 6: Publicly Release Data, Including Through a Customizable Interface
Giving data back to states and institutions cannot be the only purpose of a federal student-level system. Information in the system must also be widely available for students, families, the public, policymakers, and researchers.

Any disclosure of data in this system to external stakeholders must take into consideration the extent to which these groups really need information on individuals, versus more aggregated or summary statistics. The closer data disclosure gets to student-level results, the greater the need for substantial privacy and security protections.

Students and families. Students and families need a way to access public data that is as personalized to their situation as possible. They need an easy interface that allows them to see outcomes for students like them, by at least institution, and ideally program. For example, such a system must allow a low-income Latino student who is deciding between engineering programs at three different colleges to understand how their choices might differ in terms of their chances of graduating, the amount they might expect to borrow, their ability to repay loans, their potential earnings, and other relevant information.

Policymakers. Policymakers and the public need a way to generate aggregated statistics that can help answer key policy and college choice questions. This includes needs such as judging repayment rates over time by institution and type of college, generating completion rates for Pell students, and other analyses. The system could meet their needs by creating a tool similar to PowerStats, something that NCES created several years ago to allow the public to query its large sample surveys without risking access to individualized data. In addition, ED should make efforts to release data aggregated to the institution and program level within institutions to inform the public on results.

Researchers. Individuals conducting research in higher education need to have various ways of accessing federal data. For most researchers, a system akin to what is available for policymakers should suffice and will make it possible to answer straightforward questions and describe the state of higher education in terms of completion rates and other outcomes.

In other cases, however, researchers may need access to individual-level data. Doing so will help them construct more complex analyses, including running regressions and grouping students in different ways than standard reports might allow. In these situations, access to the student-level data should be very tightly controlled. NCES's process for allowing access to student-level data from its sample surveys provides an example of how to handle this issue. This includes requiring anyone who is going to use the data to complete a licensing agreement from NCES, promise not to disclose data under penalty of jail time or a substantial fine, and provide a security plan to keep the data safe, among other steps.[56]

Putting the Picture Together
Table 3 lays out the complete picture of what types of information would be in a student-level system as well as who would report them.

Table 4 lays out who sends what information into the system and what they get back from it.

Reduction of Other Federal Data Collections
In addition to the benefits already described above for institutions, developing a student-level data system should substantially lessen the burden that colleges face in fulfilling IPEDS reporting requirements. Of the 12 IPEDS surveys and their components, seven could be eliminated.[57] An eighth—the institutional characteristics survey—could be mostly eliminated if these data are added to the student-level system. Three of the remaining surveys ask about issues unrelated to student outcomes: finance, human resources, and academic

TABLE 3: DATA ELEMENTS IN A STUDENT-LEVEL SYSTEM

Indicator Type	Example variables	Who reports?	New reporting?
Student identifiers	Name Address Social Security number Birthdate	Institution	Only for students not receiving federal aid
Student enrollment	Program Credential level Attendance intensity Credits attempted Credits earned Credit accumulation Completion flag	Institution	Credit accumulation is new for all; other indicators are only new for students not receiving federal aid
Student non-federal aid	State grants Institutional grants Institutional loans Other non-federal loans	Institution	Yes, at the student level
Student price charged	Amount of tuition, fees, books, and supplies charged to students prior to subtracting grant aid	Institutions	Yes, though institutions need these data to populate existing net price calculators
Student demographics	Expected family contribution Dependency Income upon entering College readiness Marital status Parent information (if applicable): Income Education level	Department of Education / Institutions	College ready is new for everyone; other indicators can be pulled from FSA systems or federally aided students
Student federal aid	Grants and loan amounts, types, dates of receipt Loan status Loan default or delinquency Repayment plan Use of public service loan forgiveness	Department of Education	No, in National Student Loan Data System
Earnings	Mean and median earnings	Social Security Administration/ Department of the Treasury	Only for students not receiving federal aid
Veterans and Service members' benefits	Military status Aid amounts, types, and dates	Department of Veteran Affairs/ Department of Defense	Yes, to Department of Education

TABLE 4. PARTICIPATING ACTORS' DATA CONTRIBUTIONS AND RETURNS

Actor	Sends into the system	Receives from system
Institution	Enrollment data Non-federal financial aid Price charged Some additional student characteristics	Earnings data Financial aid performance Transfer and completion tracking Subsequent college enrollment Results disaggregated by program and student type
State	State financial aid Verify institutional data	Performance data by state and institution that are not limited by state borders, including: Earnings data Transfer and completion tracking Financial aid performance
Federal government	Federal financial aid information Earnings information Veterans and service members' benefits	Complete picture of enrollment, transfer, completion Comparable results for students who do and do not receive federal financial aid
Students	(Nothing, others report their data)	Personalized information on outcomes, by institution and program

TABLE 5. INTERACTION BETWEEN INTEGRATED POSTSECONDARY EDUCATION DATA SYSTEM AND A STUDENT-LEVEL DATA SYSTEM

IPEDS component	Average estimated hours per college	Total estimated hours to collect 2015-16	Replaced by student-level system?
Institutional characteristics	6.3	47,250	Maybe*
Completions	10.7	80,250	Yes
12-Month enrollment	5.9	44,250	Yes
Student financial aid	19.8	148,500	Yes
Graduation rates	8.3	53,120	Yes
200% Graduation rates	3.4	21,760	Yes
Admissions	3.8	9,500	No
Outcomes measures	27.7	135,730	Yes
Fall enrollment	12.8	96,000	Yes
Finance	13.5	101,250	No
Human resources	28.0	210,000	No
Academic libraries	10.5	51,450	No

*These data are necessary for both IPEDS and the student-level system but only need to be collected in one place.[59]

libraries. A student-level system would never replace these data, though it is debatable whether some of them, such as the libraries data collection, are necessary. A fourth—the admissions component—deals with students who do not enroll, and thus this data could not be replaced. However, this collection only consists of 23 indicators, NCES estimates that the fewest number of colleges need to fill it out, and it is the fastest and least burdensome survey to complete.[58]

Replacing seven IPEDS collections and most of an eighth have significant implications for burden reduction. Based upon NCES's estimates, these changes would reduce IPEDS reporting time by roughly 60 percent, saving around 88.6 hours per institution and 579,610 hours overall (not including the institutional characteristics survey).[60] The potential for increased burden from the student-level system at least during the implementation phase will offset some of these savings, but colleges should on net still see a decrease since they are already reporting much of these data to NSC.

Replacing IPEDS reporting with a student-level data system will also save time and alleviate the burden of any future variable changes for institutions. Right now, any time IPEDS changes variable definitions or adds new cohorts (such as completion rates for students receiving Pell Grants) the institution must construct new cohorts. As a result, the institution may be reporting outcomes for the same student multiple times across several cohorts. By contrast, creating a new completion measure in a student-level system would likely just mean changing the database's programming to cut the data a different way, since the necessary base information is already

included. This not only means institutions will not have to report more data, but it will also be easier to create retroactive measures that look back in time to measure results. This is not available in IPEDS, where it would be infeasible to ask institutions to go back and report years of data for a new measure.

The interaction between the student-level system and IPEDS also matters from an institutional perspective. That is because at many colleges the people who handle the reporting for NSLDS and student financial aid purposes and those who do it for IPEDS are different. In general, institutional research offices most likely do IPEDS reporting, and in 34 cases, a state or university system either submits or verifies IPEDS data. By contrast, the registrar, the financial aid office, or both handle the reporting into NSLDS or NSC. This means that the IPEDS data undergoes a different degree of quality control checks by trained researchers than transactional data for financial aid purposes. Moving to a system by which the data sent to FSA becomes the basis of generating outcomes information will mean that colleges will need to involve their institutional research offices more in these processes. It might also mean rethinking university structures to ensure that the institutional research, registrar, and financial aid offices are better aligned. These types of changes could have spillover benefits, such as helping a college better manage data requests and issues in general.

Finally, a student-level data system would also reduce the amount of information that the NCES sample surveys currently collect. For example, questions about where students are enrolled and if they finished, which are included in the Baccalaureate and Beyond Longitudinal Study would no longer be necessary because the data system could answer them. This does not mean the surveys should be eliminated. There are still valuable questions they ask that a student-level system would not capture, such as the classes students took, the type of jobs they held, how many hours a week they worked, and if they received aid based on need or merit. It also includes items regarding students' revealed preferences learned through interviews, which encompasses questions detailing a student's ultimate intentions, whether student debt affected their behavior, or why they dropped out.

INTERACTION WITH THE NATIONAL STUDENT CLEARINGHOUSE

This design for a student-level data system does not presume requiring or eliminating the National Student Clearinghouse. Rather, it assumes colleges could continue to use it if they wish or cease using it if the functionality it provides is no longer necessary. To that end, this system must have the capability of accepting data from NSC, much like how NSLDS currently does. This will allow institutions to handle their student reporting through NSC if they desire. The system, however, must also have its own process for institutions to report data if they do not want to use NSC.

Privacy and Security

Privacy and security must be a constant consideration for everyone designing, reporting to, and using the student-level data system. Achieving these goals is not just about establishing hard and fast rules, although criminal liabilities already exist and should be continued for breaches in privacy and security. Rather, privacy and security are also about adopting a mindset continuously attuned to vulnerabilities.

The exact way to ensure the privacy and security of the student-level system will ultimately be something that its builders and users must determine. Nevertheless, they should take into account several guidelines or approaches. While this paper briefly summarizes these ideas, please see Joanna Lyn Grama's paper, *Understanding Information Security and Privacy in Postsecondary Education Data Systems* for a thorough discussion.

In general, good privacy and security protocols rest on three foundational concepts, *confidentiality*, *integrity*, and *availability*, or the CIA security triad.[61]

Confidentiality

Confidentiality is ultimately about putting protections in place that prevent unauthorized access to data. This can come in many forms. The most obvious is ensuring that the database is secure enough to prevent people who have no business retrieving or using the data from accessing it. Of course, this includes minimizing opportunities for hacking or data theft. While confidentiality is a crucial issue, far more detailed bodies of work on this topic already exist. This includes the robust data privacy and security protocols of the National Institute of Standards and Technology (NIST).[62]

Confidentiality is also about ensuring that the roles for everyone who could access the data are properly defined and that people cannot access sensitive data unless doing so is absolutely necessary. In addition, the system designers and managers must tightly control who determines what is necessary. Addressing confidentiality also requires careful determinations of what types of data within the system can be accessed and for what purpose. This includes ensuring that authorized users can only access data for specific purposes related to their role. Such considerations need to be defined not just for data overall but also on an indicator-by-indicator basis.

Integrity

Integrity concerns not only the processes around how the system ensures data are correct, but also who has the ability to edit or fix data. This too gets back to questions of who should have what type of access to the data. For instance, an institution should be able to edit or correct data that it submitted about their students. Yet it should only be able to edit data connected to a time when that student was part of their institution.

In other words, an institution should not be able to alter data on the price charged to a student at a different school or update the completion indicator that signals that one of its students graduated from a different institution. From a student perspective, the Family Educational Rights and Privacy Act (FERPA) has already laid out some requirements and guidelines for how individuals can inspect or correct their records.[63]

Availability

Availability has some overlap with confidentiality and integrity, but it also addresses some technical aspects. Availability means the data system's design offers sufficient backups and redundancies to ensure that it will be accessible when people need it. This means authorized users seeking to engage with the system for an allowable purpose should be able to do so in a reasonable amount of time. This concept also requires making data available to people who might need to engage with it in different ways. For instance, addressing data availability for the public in this system looks quite different from how it would for institutions. In the public's case, availability means a tool for generating aggregated reports that can tell stories about trends in the data or results across institutions, within a program in a given school, as well as national stories by type of institution or state. By contrast, availability for institutional users means an easily navigable system that does not crash and allows them to upload or download data with relative simplicity.

Table 6 lays out some ways to think about possible policies for confidentiality, integrity, and availability for different types of users. Within each user type, access would be limited based on individuals' roles. For example, at an institution, only those who currently have access to personally identifiable information (e.g., a financial aid officer) would be able to access individual-level records. Similarly, the federal government user here would be someone engaged in the administration of the federal student aid programs, not just any federal employee.

Understanding how NSLDS fares under the CIA triad may be a useful way to further consider how a student-level system might work. Right now, NSLDS arguably does quite well on some forms of confidentiality and integrity, but still needs work. In terms of confidentiality, it is arguably too restrictive. On the one hand, NSLDS tightly controls access to the database in a way that makes it difficult for completely unauthorized users to get into the system. Yet it has struggled in the past with authorized users engaging with the database for unallowable purposes. For instance, in the mid-2000s, student loan companies were caught using NSLDS to identify students for marketing purposes—a goal that is far afield from the reason why they had access, namely to update information about borrowers while in repayment. While ED fixed that issue, it has had other troubles with federal employees with authorized access using the system for improper purposes, particularly looking up the student loan records of President Obama. This

TABLE 6. A HYPOTHETICAL APPLICATION OF THE CIA TRIAD TO THE STUDENT-LEVEL DATA SYSTEM

User type	Type of student data visible	Confidentiality (type of access)	Integrity (extent to which they can edit data)	Availability
Institutional representative working directly with student data	Personally identifiable	Student-level data of people who are attending or have attended that school	Edit data related to the time that student spent at that school (e.g. cannot change completion status at another college)	Ability to access allowable data on demand with minimal to no slowdowns
State representative working directly with student data	Personally identifiable	Student-level data of people who are or have been enrolled in public colleges in that state	Edit data related to the time that student spent at that school in their state	
Federal government engaged in active management of aid programs	Personally identifiable	Student-level data of people who are or have been enrolled in college	Data related to information related to the federal governments' interaction with that student (e.g., federal aid amounts)	
Other federal agencies uploading data	Personally identifiable	Information only needed for purposes of matching data	Data they are matching into the system.	Ability to access data on students who match their records
Other federal users	Treated like researcher or public, depending on use	Treated like researcher or public, depending on use	None	Public tool for aggregated analysis
Researcher	Depends on access protocols, at least de-identified	Student-level data of a representative sample of everyone in college	None	
Public/Policymakers	Anonymized or aggregated	Aggregated student data of people who have been enrolled in a given college or program	None	
Students	Personally identifiable	Their own record	Edit data stemming from student reporting, such as parents' marital status	

suggests a need to stay vigilant about ensuring that authorized users are only using the system for proper purposes.

At the same time, NSLDS arguably goes too far in terms of confidentiality for other types of users. Right now, FSA does not provide researchers any way of accessing student-level information stored in the system, even if they follow an intensive privacy protocol. Researchers frequently have legitimate reasons for wanting this type of access in order to evaluate the programs, which the current restriction hinders. Similarly, the public only recently gained aggregated data access to the system through the release of institutional-level data on the College Scorecard. This is an important step to ensuring that ED strikes the proper balance between confidentiality on the student-level data while allowing broader access to de-identified or aggregated information.

The other place where NSLDS could likely improve is on availability. Given the age of the system, it can be very clunky and difficult to extract data. Moreover, FSA staff have said anecdotally that any sizeable information request from the system frequently runs the risk of slowing down the ability of users to engage with the system for operational purposes. If this happens, it can be harder for schools to provide federal aid funds to students when they need them. Improving on this challenge will be an important step to ensure that NSLDS is sufficiently available for both operational and evaluation purposes.

Options for Creating a Student-Level Data System

Unlike other potential improvements to the national postsecondary data infrastructure, a comprehensive and integrated student-level system does not exist at the federal level. Establishing one will not be without significant difficulties, but it is most certainly feasible. The current statutory ban on creating it is the most obvious roadblock, but even a change in the law will not solve other issues related to location, capacity, timing, and cost.

Setting aside the political feasibility of overturning or creating an exception to the ban, this section considers four options for the location of this new database. A new student-level database could be built as (1) an expansion of NSLDS, (2) a new data system in FSA, (3) a new data system in NCES, or (4) a hybrid option in which FSA handles data collection and NCES handles public data analysis and access. All cases assume the student-level system would replace most of IPEDS, but given the shift from an institutional system to a student one, it may be easier to think of the third option as a new database. The following paragraphs discuss the implications of choosing any of these three options in terms of implementation challenges, timing, cost, and other major issues. **Table 7** provides a simplified breakdown of the pros and cons for each of these options.

TABLE 7. FOUR OPTIONS FOR BUILDING A STUDENT-LEVEL SYSTEM

Description	Pros	Cons
Expand National Student Loan Data System (NSLDS)	Cheapest/fastest Uses existing reporting streams Minimizes duplicative reporting	Clunky/outdated Hard to access Conflicts with FSA operations Needs improvement in transparency and analytics
Create new Federal Student Aid (FSA) database	Modern, up-to-date Leverages FSA flexible contracting Well resourced	Expensive/slower to build Struggles with transparency and analytics Could conflict with FSA operations
Create new National Center for Education Statistics (NCES) database	Maintains good track record on transparency, privacy, data standards Separate from Title IV operations	Expensive/slower to build Fewer staff and resources
Use FSA for data collection, NCES for public dissemination	Uses FSA data collection strength and avoids weakness on transparency Uses NCES data dissemination strength, avoids lack of resources and collection capacity Separates state and institutional access from public and researchers Would cost less than brand new database	Requires establishing NCES and FSA data sharing agreements Needs to work out proper timing of sharing and data corrections process Offers two sets of privacy and security protocols

Option 1: Expand NSLDS

NSLDS already stores almost all of the information needed to establish a comprehensive student-level system. For instance, of the metrics discussed earlier, the only ones not already included in NSLDS in some form are a flag for college readiness, the student's race/ethnicity, the student's level in college (which is stored for borrowers in a different FSA database), non-federal financial aid, and data on credit accumulation. Everything else—a student's program, age, etc.—is already collected as part of the operation of the federal aid programs.

What NSLDS lacks is completeness. It only includes data on students who received federal aid, and it only updates information on still-enrolled students in terms of when they receive federal aid. In other words, an enrolled student who received aid in the past but not in the current term may not have all of the corresponding information about programs and courses that an aided student does.

In most cases, changing NSLDS to collect the additional necessary data indicators as well as store information on all students in college should be quite straightforward. That is because NSC already contains these data for most colleges. For instance, NSC already asks participating institutions to report data on all their students to help ensure accurate coding of borrowers' enrollment status. It also has optional flags for race/ethnicity and remedial status. Credit accumulation is the only necessary measure not currently captured by NSC, though it is working to collect this on a voluntary basis.[64]

Given these considerations, expanding NSLDS to store information on students not receiving federal aid should only require some programming changes to account for additional fields, an increase in records, and other technical fixes. This is certainly doable. For one, NSLDS has already seen a significant increase in the number of records and amount of information in a short period. From February 2010 to July 2013, the number of students in NSLDS increased by 17 percent, representing 12 million additional individuals.[65] Second, NSLDS already has stored data on some un-aided students for a brief time as part of a regulation that held career-training programs accountable for the debt levels of their graduates. This rule, known as gainful employment, originally required institutions to report data on all students enrolled in these programs, even if they did not receive federal aid. These data were stored in NSLDS. However, after a successful legal challenge from the main trade association representing private for-profit colleges, a judge ruled that collecting these data violated the student-level data ban and required their deletion.[66]

Even if the programming changes that turn NSLDS into a comprehensive student-level data system are straightforward, cost and time considerations still matter. Gauging these elements can be very difficult since the federal government does not disclose pricing terms for individual components of contracts. Fortunately, the government does release contract modification data on USASpending.gov. These data provide some guidance for estimating potential costs. Based upon a review of the more than 100 listed transactions on the NSLDS contract since 2010, ED pays several hundred thousand dollars to add indicators to the system, with the price varying based upon the complexity of the change. For instance, ED paid $210,000 in March 2014 to update the repayment plan schema in NSLDS.[67] By contrast, the implementation of the

more complicated requirement that students cannot receive subsidized loans for more than 150 percent of their program length cost over $1.6 million.[68] Given this range, simpler data like a flag for college readiness or for race/ethnicity and a count of credit accumulation should be on the lower end of costs.

Expanding the universe of students covered would also come with costs. Based upon prior spending data, the programming changes to accept these students would probably cost around a million dollars.[69] This estimate comes from looking at some of the expenses related to gainful employment reporting in 2012, which expanded the individuals captured as well as added new variables.[70] Beyond programming, FSA would also have costs in terms of training institutions to report the additional data, and for careful review of results for at least the first year or two to ensure colleges enter information correctly, etc. Most of these costs could presumably be borne by the existing staffing infrastructure and events like the annual FSA Conference.

Matching NSLDS data with information from other agencies would also require resources in terms of personnel, time, and programming. Existing budgets could cover these first two areas. ED staff and other agencies would have to collaborate to establish memoranda of understanding that would allow for data sharing. Any agency involved may also have to alter the documents that spell out the uses and allowable disclosures of their data systems, commonly known as a "system of records notice." Once the legal procedures are in place to allow the data to flow, ED would also have to cover some programming and data matching costs. For instance, when ED worked with SSA to establish the match that allowed it to obtain earnings data on graduates in career training programs, it had to pay SSA roughly $250,000 to create new queries. In general, these programming expenses should be infrequent if not one-time costs, as the variables should not change. Finally, ED would have to pay a fee to the other agency for each record it matches. This expense is typically very low. For instance, in the case of the SSA match, ED had to pay a tiny fraction of a penny for each record it matched. These would be ongoing costs for each new query, but again, they are quite modest sums in total.

The other advantage of making NSLDS a complete student-level system is that FSA could expand its capabilities through a modification of the existing vendor contract. This would allow for making changes quickly. Add in training time, the establishment of data definitions for the college readiness flag and credit accumulation measures, any additional memoranda of understanding between agencies, as well as test data collections, and the entire system could likely be in working form within one to two years. Simply adding information on un-aided students without the new indicators could

happen even sooner since it would only require adding data on students that it previously had excluded.

While using NSLDS for a student-level system has much appeal, this option also contains substantial drawbacks. Some of these are technical, but others are cultural and just as challenging since they do not present an easy fix.

From a cultural standpoint, locating the student data system within FSA could be problematic. The largest challenge is that FSA has not always prioritized transparency, research, or analysis among its core functions. Instead, it has typically focused on the administration and operation of the financial aid programs, with its goals and processes oriented toward those ends. As a result, it may not have all the necessary staff and procedures in place to run a student-level system with a more outward facing component. For example, it would need to create a process for working with the higher education community to define key metrics the way NCES does. It would also need to employ more statisticians with research backgrounds who can handle technical standards and other considerations necessary for figuring out appropriate rules for issues such as privacy suppression of small cohorts. Finally, FSA would have to place a greater emphasis on data transparency.

To its credit, FSA is trying to improve on a number of these issues. Until the beginning of the Obama administration, FSA released little data about the student aid programs except for student default rates on federal loans by institution. Now, it operates a Federal Student Aid Data Center, which includes quarterly updates on the amount of aid dollars flowing to institutions by program as well as overall updates on the state of the entire loan portfolio. It also went even further with data release on the College Scorecard; for the first time it disclosed institution-level information from its data systems about completion, loan repayment, aggregate debt, and earnings. Even better, FSA disaggregated these data by different types of student characteristics such as dependency status, Pell Grant receipt, gender, and income upon entry.

There is hope that the agency may continue to improve on the transparency front. It has spent the past several years building an enterprise data warehouse that stores exports from NSLDS, which may make that data more accessible without affecting operational needs. If a new database ends up being the best option for the student-level system, the warehouse could also provide the necessary financial aid data for matching. The status of the data warehouse, however, is unclear and FSA has not disclosed who will ultimately be able to access it.

Even still, FSA lags behind other federal agencies in terms of allowing researchers access to its data and systems, even in tightly controlled situations. For example, the Internal Reve-

nue Service, an agency that protects data that are arguably more important than what FSA holds, has a policy in place that allows researchers to gain direct access to records. This IRS process is not easy to navigate and has very strict rules that include requiring the researchers to physically go to the agency's offices, obtain approval for any papers discussing their analysis, and submitting formal research proposals.[71] FSA has no similar procedures in place today for research access to data.

Beyond FSA cultural issues, NSLDS also has significant technical limitations. For one, NSLDS is widely recognized as outdated, clunky, and not particularly user friendly. The trade association that represents financial aid administrators has referred to NSLDS as "notoriously and unacceptably slow."[72] Even screenshots of the system show an interface that resembles Internet web pages from the late 1990s.[73]

Source: NSLDS Enrollment Reporting Guide p. 59.

Fixing NSLDS's problems will be difficult. Like many government databases, NSLDS's programming largely uses COBOL, a very old coding language that only one-quarter of U.S. colleges teach, with an even smaller share requiring computer science graduates to learn it.[74] This limits the number of vendors who could improve or maintain the system. Even worse, FSA designates the competition for the NSLDS contract as a small business set-aside and continues to award it to Briefcase Systems Development, a company that appears to operate out of a house in Great Falls, Virginia and has no web presence.[75] FSA also re-competed this contract in 2014, meaning it would have to cancel future option years to get a new vendor.[76]

Adding millions of additional records to a database that already has a history of poor processing speed may be a particularly bad idea since NSLDS must continue to perform its other operational roles. The system is crucial for ensuring that students receive the grants and loans they need and for

protecting against waste, fraud, and abuse. To that end, further slowdowns due to the records of students who are not in the aid programs could negatively affect FSA's operational performance.

Given the significant concerns about FSA providing sufficient data disclosure and access, any legislation that creates a student-level data system must contain provisions to ensure that the database is usable and accessible with proper privacy and security protections. This should include mandates for data dissemination by FSA, a researcher license process, and a public data analysis tool. Policymakers should also update the structure of FSA's performance-based organization to make data analysis and disclosure explicit, required steps. Doing so emphasizes that data dissemination must be a core part of FSA's mission. These elements should also have explicit goals attached to them in FSA's strategic plan and the contract for its chief operating officer. This will provide a hook for the Secretary of Education to hold FSA accountable for properly operating a student-level data system.

It is unclear whether the drawbacks discussed above make NSLDS an infeasible location for the student-level data system. Theoretically, the lower startup expenses could allow for additional spending on NSLDS modernization, which would have additional benefits for overall improvement in the operation of the federal aid programs. Nevertheless, simply changing NSLDS without addressing any of these other factors could lead to a student-level system that is insufficiently accessible for the needed purposes.

Option 2: Create a New Database within FSA
Establishing the student-level system as a new database that operates within FSA could correct for many of the potential flaws that come with expanding NSLDS. Starting from scratch would make it possible to build a new system that incorporates the latest in database technology and with an eye toward making it adaptable to future changes. It would also allow for discussions with end users to ensure that it is easy for schools and others to use, including built-in public and institutional data analysis functionality.

Locating a new data system within FSA brings a couple of unique benefits. For one, the office has a special designation as a performance-based organization.[77] This status affords FSA special flexibilities in terms of hiring and contracting that allow it to pay talented staff more, requisition vendors with an eye on achieving results, and establish other processes that could help put the resources in place necessary to establish a new data system. FSA is also the single largest office within ED, totaling 1,300 employees.[78] This gives it by far the most staffing resources to put toward assisting colleges with the transition to a new system without needing a significant increase in the number of employees.

Building something new would cost money and time. Currently, FSA spends about $8 to $11 million a year just for maintaining NSLDS.[79] Given programming costs of a few hundred thousand dollars for each change to the system, this suggests that creating a new student-level database would be more expensive than expanding existing systems. Whatever the cost, however, it would be a tiny sum compared to the roughly $130 billion disbursed through ED's financial aid programs each year. A new system would also take time to build. In 2006, ED informed the Office of Management and Budget that planned re-engineering of NSLDS would take at least three years.[80] Given advancements in software development over the last 10 years, this process would likely be faster, but it would still be a minimum of several years to create a new system.

Developing a new system would also mean navigating a transition period. Until the new system is capable of replacing the enrollment reporting that currently goes into NSLDS, ED would need to find a way to populate both systems without colleges having to report twice. This could mean possibly a temporary expansion of NSLDS that then exports data to the new system or perhaps a test process that would permit a set of institutions to try out the new reporting before expanding it to everyone else. Finally, the new system would need to develop linkages with all the necessary databases. This includes pairing with NSLDS, SSA, IPEDS (for institutional information), COD, and CPS. Doing so should not be too difficult for the databases within ED, but it could mean signing new memoranda of understanding between agencies and possibly some programming changes to reflect the new path for data flow.

Importantly, FSA would still have to update its mission, values, and capacity to meet increasingly complex demands of its growing body of stakeholders. If the new system did not come with changes to how the agency approaches information release and analysis, then the data could end up locked away in a system that is modern and flexible, but still unavailable.

To be clear, these downsides do not make the idea of a new system prohibitive. The benefits of a modernized system are significant. However, pursuing this option would be a multi-year endeavor. It would require building new systems from scratch, training every single institution to use it, and establishing an overlap period to work out kinks. This period will be unavoidably rough, though the product would almost certainly be better than just expanding NSLDS.

Option 3: Establish a New Student-Level System within NCES

If the concerns about FSA's culture and willingness to grant access to data are too great to overcome, then NCES may be a better option for housing the student-level system. While a substantial overhaul of IPEDS could potentially create this data system, it would change so much that it is easier to think of this option as building an entirely new database.

Locating a new student-level data system within NCES presents several advantages. Unlike FSA, NCES does not have to balance operating the federal financial aid programs with demands for research and analysis. It already has extensive experience with student-level data from large sample surveys and tools in place to allow researcher and public access to data it holds. NCES also has rigorous privacy protocols for protecting student information, including severe penalties for misuse of data, such as felony charges, fines of up to $250,000, and as much as five years in prison.[81] A student-level data system operated in NCES is thus likely to be more open and accessible—while still operating through an existing and tested protocol for privacy protection—than one situated in FSA.

NCES's other major strength is that it already has processes for establishing variable definitions. Through its operation of IPEDS, the agency solicits expert opinion to properly define and refine measures. For instance, NCES convened expert stakeholders for detailed discussions about the content of the college ratings system, which later became the revised College Scorecard.[82] This means NCES staff could work with the higher education community to define new variable indicators such as credit accumulation and the flag for college readiness. Regardless of whether NCES is the ultimate host of a student-level data system, it should still hold expert meetings to inform the data indicators included.

From a funding standpoint, NCES is the office most likely to realize savings from the creation of a student-level system, since it would replace over half of the reporting that goes into IPEDS. It could also fill in some of the data currently collected on the sample surveys, freeing up resources to ask additional questions or possibly expand administrations. This would make it easier to use funds saved on IPEDS to help pay for the ongoing operation of a student-level database. Finally, because NCES is a research organization, having it control the student-level data system would send a stronger message about using the data for analytical purposes, potentially reducing some of the concerns about misuse.

Establishing a student-level data system within NCES would not diminish any of the difficulties associated with creating an entirely new database. All of the cost, timing, and operational considerations related to setting up a new system within FSA apply here. NCES would also have some unique difficulties. For one, the office is much smaller than FSA, with only about 115 employees conducting work across all levels of education.[83] Consequently, any new data system would likely require an increase in staffing.

Operating a student-level system that also has an operational role in the federal aid programs could place NCES in an awkward position. As part of the Institute of Education Sciences, NCES is independent from the rest of ED.[84] Operating a student-level data system would require establishing operational linkages with FSA to fulfill NSLDS reporting requirements. This could place the agency in a strange position of needing to occupy a less independent role with respect to the aid programs.

Ultimately, the selection of NCES over FSA for a new data system would reflect the importance of transparency and disclosure. Getting data into the hands of the public and researchers is one of NCES's strengths. Locating it within a statistical agency would also make the student-level system clearly about fulfilling the functions provided by IPEDS, which are less politically controversial. Relying on NCES for a student-level data system, however, is a larger stretch from an operational standpoint due to staffing considerations and its traditionally independent role. To answer whether these concerns are prohibitive, policymakers should consider how they rank analysis, burden, and transparency in the effort to set up a student-level system.

Option 4: A Hybrid Reporting Option
Making a student-level data system work requires getting the information collection right while also having an orientation that allows for careful access to data at differing levels of aggregation. Unfortunately, the three options above suggest that neither of the two logical places to house a student-level data system can perform both functions exceptionally well. FSA already has the widespread collection of student-level data in place, and it has the direct operational linkages necessary to ensure the new system does not add too much new burden. However, FSA also has a mixed track record on data release and analysis. NCES, meanwhile, excels at multiple types of data transparency, including aggregated data and carefully controlled access to student-level information. It, however, has no experience with collecting information on millions of students and no formal role in federal student aid operations.

This final option therefore considers a hybrid approach designed to utilize the strengths of each office. Under this setup, FSA would handle the actual work of data collection while NCES would take charge of dissemination and researcher access.

Here is a hypothetical way this option could work: FSA expands NSLDS or uses the Enterprise Data Warehouse so that institutions start submitting the necessary enrollment, demographic, non-federal financial aid, and pricing data on all students. It also establishes new memoranda of understanding to allow for more immediate reporting of data from the Veterans Administration and Department of Defense on stu-

dents served by those agencies. This creates a comprehensive database with most necessary variables. Once a year FSA then sends data from NSLDS to the Social Security Administration to generate mean and median earnings figures. This completes the rest of the data needed for the system. For any new data collection or calculation, FSA would rely on NCES and a technical review panel to establish definitions.

States and institutions seeking access to these data would continue to do so through FSA. Perhaps by using the Enterprise Data Warehouse, FSA could create a more modern interface that has a few new reports. Furthermore, access to the student-level data would continue to occur through roughly the same protocols that govern NSLDS and other FSA systems access today (with any necessary security updates).

Meanwhile, NCES would handle consumer, policymaker, and researcher access to a student-level system it administers. At least once a year, FSA would commit to transmit a data extract to NCES. Before doing so, FSA would allow colleges to review and correct any errors in their data. Once the data go to NCES, it would become responsible for generating public reports of various outcomes by institution or program, populating what schools would have normally reported to IPEDS, the College Scorecard, or other data entities. NCES would also handle access protocols, privacy rules, and other technical considerations for researchers and the public.

This setup provides several advantages. First, it clearly splits the responsibility for user access to the system so that FSA can keep focusing on operational concerns while allowing NCES to continue handling the researcher agreements and public transparency it already does. Second, NCES's new data system would not need to be as complicated since it would only receive an export from FSA and not build any linkages itself. Third, schools would be able to correct data and still get access to it more immediately to fit their needs, ideally reducing the rate at which inaccurate information becomes public. Finally, FSA would not have any formal expectations for transparency, reporting, or analysis, freeing it to continue to address operational issues.

Making this setup work requires sorting out a few issues related to timing. FSA and NCES would need to figure out a proper schedule for how often to transmit data. This would include establishing policies for when to send updates if an institution corrects data. Ideally, building in a lag time between information collection and transmission to NCES (such as several months after each federal aid award year ends) would cut down on corrections, but the possibility still exists. Similarly, NCES would have to make sure the database is capable of storing data for multiple years so that longitudinal assessment is possible.

Admittedly, this option is somewhat more complicated than the others mentioned above. It would mean that the student-level system exists in two places, not just one, which in turn would mean two sets of privacy protocols, security protections, terms of user access, etc. This duality would not only require close collaboration between NCES and FSA to ensure the process goes smoothly, but it would also mean both agencies would need to contribute to the startup expenses.

Despite all these considerations, this option represents a best-of-both-worlds approach to the challenges of data collection and dissemination. It allows the two offices in ED best suited to the student-level system to do what they do best and acknowledges that there is a meaningful distinction between the purposes of the system for states and institutions versus the public, students, and researchers. It would undoubtedly require a bit more coordination, but if done well it should result in a system that works well for all stakeholders.

Recommendations for Institutions
Improve Alignment between Institutional Research and Financial Aid Reporting
This paper strongly recommends establishing the student-level data system through an expansion of the existing individual-level reporting that already occurs for the purposes of the federal financial aid programs. This presents the best path for balancing the goals of comprehensive data coverage while minimizing additional work for institutions.

Doing so, however, could require some changes for institutions. That is because this system would start to calculate outcomes from transactional data that either registrars or the financial aid office currently report. At most colleges these offices are likely distinct from the ones responsible for generating the outcomes data that are currently used in IPEDS and other public tools. This matters because the two offices may have different protocols for cleaning up data, with the institutional research office having more experience in this space.

There is no clear federal solution to this problem. Rather, institutions may need to investigate their own internal structures to ensure strong partnerships exist between institutional research offices and others who are reporting student-level data. This may include moving these entities under the same vice president or provost or other changes that foster stronger communication. Only by better aligning these offices can the college have greater confidence that the data used to generate outcomes are accurate.

Recommendations for Federal Policymakers
End the Ban on a Student-Level System
Allowing the federal government to collect comprehensive information on all individuals in higher education will provide significant benefits to institutions, policymakers, students, and the public. The current legal provision preventing the creation of such a system results in undue burden for institutions in terms of data collection, presents incomplete pictures of colleges' success and return on taxpayer investment, and makes it harder for consumers to choose among institutions. The wide support for ending the ban from members of both political parties, as well as several associations representing institutions, shows that it is time for Congress to overturn the ban.

Add Data Dissemination and Disclosure to Federal Student Aid's Organizational Goals
The advantage of being a performance-based organization is that FSA is accountable for achieving certain outcomes. Currently, the main goals set for FSA in statute relate to operational efficiency, such as sending dollars to students quickly or reducing data duplication. These are important goals, but proper management of the aid programs must be about more than just the flow of dollars. The sheer size of the federal aid programs demands robust analysis and transparency for the public about how loans and grants are performing. FSA is making strides in this area, but it has room to improve.

Adjusting the terms of FSA's performance-based organization to emphasize data collection and dissemination could help the agency do more in this area. Making data analysis and transparency explicit requirements will ensure that the agency and its chief operating officer are evaluated on their success in these areas and will place greater pressure on them to perform well. This change would improve the usability of FSA data even in the absence of a student-level system.

Modernize the National Student Loan Data System
Regardless of whether it becomes the ultimate home for a student-level data system, NSLDS is too important a database to continue entrusting it to a small business contractor using an outdated programming language. It needs a complete rebuild with sufficient resources that reflects the technological, privacy, and security needs of today. FSA has two options for fixing this problem. First, it could hold a new competition for a contract to replace NSLDS while keeping the existing one in place until it expires. Alternatively, it could cancel the remaining option years on the existing NSLDS contract and hold a new competition that does not restrict the winner to being a small business. Larger vendors might have more resources to fix the system and may be able to update the system at a lower cost.

Create a Researcher License for the National Student Loan Data System

NSLDS must become more accessible to researchers. The sheer amount of taxpayer money spent in federal aid programs demands more transparency into their outcomes and efficiency. Other federal agencies protecting data that are arguably even more sensitive have created processes for qualified researcher access. As discussed earlier, the Internal Revenue Service recently allowed economists Raj Chetty and Emmanuel Saez to conduct a groundbreaking study using tax records.[85] It did so through a request for proposals process that FSA could replicate. Exploring such an option at FSA, perhaps assisted by NCES to help with vetting and judging proposals, could go a long way toward opening up these data.

Convene Technical Review Panels for Necessary New Indicators

While the current NSLDS structure includes the most important indicators necessary for a student-level system, there are a few missing data points that should ideally be included. If Congress overturns the ban on the student-level system, a qualified group of stakeholders will need to define these new indicators.

Fortunately, NCES could start the process of working with the postsecondary data community to define these measures now. It should start by convening technical review panels to develop the proper definitions of a college readiness flag, credit accumulation variables, non-federal financial aid, pricing information, and other necessary indicators. This approach would incorporate the postsecondary community's feedback now and reduce startup time if Congress overturns the ban.

It may seem strange for NCES to start defining variables before a student-level system exists, but making progress on these indicators would be worthwhile now. This is also in keeping with the existing ED-driven Common Education Data Standards movement, which tries to create voluntary uniform data definitions for everyone in education to use.[86] At the very worst, NCES could use these discussions to develop new measures that could be added into IPEDS, since it has the ability to make changes to this collection without congressional approval. For instance, the work on a college ready flag could become a new variable for future IPEDS collections. The advantage of this approach is that colleges will also be used to reporting these data at the institutional level, so adjusting them to a student-level collection will not be as difficult.

Reconsider Privacy Rules on Earnings Disclosure

Americans view their earnings information as deeply personal and private. That is why it is understandable that disclosure rules from SSA and Treasury implement significant restrictions on how other agencies receive these data. Nevertheless, the current rule requiring the aggregation of earnings information up to at least 10 people with at least 10 changes across cohorts hinders data quality. In particular, it makes it more difficult to generate income averages across multiple disaggregated cohorts.

A few tweaks to the earnings disclosure rules may help balance the need for privacy with greater assistance for research and analysis. Treasury and SSA could change their rules to allow for re-running income data for similar groups that have fewer than 10 different individuals in them. While these current restrictions theoretically make it possible to avoid figuring out an individual's earnings by simply generating earnings averages on two very similar groupings, a lower number of changes, such as five, could still provide such protection. This lower number of changes may make it easier to run more disaggregated earnings cohorts at a given institution.

Conclusion

Creating a student-level data system has been a politically controversial idea ever since the Bush administration seriously discussed it in the mid-2000s. Since that time, the need for one has only become clearer. In 2005, ED spent just $73 billion on the aid programs.[87] Today it spends around $130 billion.[88] The benefits of a student-level system are not just about the federal government. With states still pulling funding from their public institutions, colleges face increasing pressure to demonstrate their worth and advocate for additional money—efforts that better data could help. The process of picking a college is also an increasingly high-stakes decision for students. Selecting the right major may mean a difference of thousands of dollars in earnings each year while choosing the wrong school may negatively affect a student's odds of graduating or may lead to the accumulation of unnecessarily high debt. The increasing use of loans only raises the importance of getting these choices right. Knowing the answers to all of these questions—are colleges succeeding, where should students go, what should they study, and how should we change the aid programs to better work for today's enrollees?—are simply too important to continue trying to find them with imperfect and limited data. It is time to build a federal student-level system.

Endnotes

1   White House Office of the Press Secretary (2015, September 12). Weekly address: A new College Scorecard. Retrieved from https://www.whitehouse.gov/the-press-office/2015/09/12/weekly-address-new-college-scorecard

2   The number of colleges includes both main and branch campuses. Count of indicators listed on the following spreadsheet available within the all data download for the College Scorecard: "U.S. Department of Education (2016). "Data_File_Cohort_Map." Retrieved from https://collegescorecard.ed.gov/data/

3   Database of student information prohibited. 20 U.S.C. 1015c (2008). Retrieved from https://www.law.cornell.edu/uscode/text/20/1015c

4   Author analysis of data for the 2014-15 academic year from Federal Student Aid (2015). Title IV program volume reports. Retrieved from https://studentaid.ed.gov/sa/about/data-center/student/title-iv.

5   Author analysis of figures from National Center for Education Statistics (2016). *Integrated Postsecondary Education Data System (IPEDS) 2016-19 and supporting statement part A*. (p. 27). Retrieved from https://www.regulations.gov/contentStreamer?documentId=ED-2016-ICCD-0020-0005&attachmentNumber=1&disposition=attachment&contentType=pdf

6   Mervis, J. (May 2014). How two economists got direct access to IRS tax records. Science Magazine. Retrieved from http://news.sciencemag.org/2014/05/how-two-economists-got-direct-access-irs-tax-records

7   White House Office of the Press Secretary (2015, September 12). Weekly address: A new College Scorecard. Retrieved from https://www.whitehouse.gov/the-press-office/2015/09/12/weekly-address-new-college-scorecard

8   The number of colleges includes both main and branch campuses. Count of indicators listed on the following spreadsheet available within the all data download for the College Scorecard: "U.S. Department of Education (2016). "Data_File_Cohort_Map." Retrieved from https://collegescorecard.ed.gov/data/

9   Database of student information prohibited. 20 U.S.C. 1015c (2008). Retrieved from https://www.law.cornell.edu/uscode/text/20/1015c

10 Office of Governor Rick Scott (2015). Governor Rick Scott issues 'Ready, Set, Work' university challenge. Retrieved from http://www.flgov.com/2015/12/02/governor-rick-scott-issues-ready-set-work-university-challenge/

11 Author analysis of data for the 2014-15 academic year from Federal Student Aid (2015). Title IV program volume reports. Retrieved from https://studentaid.ed.gov/sa/about/data-center/student/title-iv. Figures for campus-based aid are from the 2013-14 academic year.

12 Federal Student Aid. (2015). Direct loan and federal family education loan portfolio by loan status. Retrieved from https://studentaid.ed.gov/sa/sites/default/files/fsawg/datacenter/library/PortfoliobyLoanStatus.xls

13 Federal Business Opportunities (2014). Attachment 1: NSLDS statement of objectives—amended. Retrieved from https://www.fbo.gov/utils/view?id=05dfcf7f45701448f3e869ea7da314f5

14 U.S. Department of Education (2013, June 28). NSLDS system of records. *Federal Register, 78* (125). Retrieved from https://www.ifap.ed.gov/fregisters/FR062813NSLDSSystemsofRecords.html

15 U.S. Department of Education (2014, January 27). Final regulations: Direct loans –150% subsidized loan limit. *Federal Register, 79* (12). Retrieved from http://ifap.ed.gov/fregisters/FR011714.html

16 Executive Office of the President. (2015). *Using federal data to measure and improve the performance of U.S. institutions of higher education* (p. 24). Retrieved from https://collegescorecard.ed.gov/assets/UsingFederalDataToMeasureAndImprovePerformance.pdf

17 National Student Clearinghouse. Transmission schedule. Retrieved from http://www.studentclearinghouse.org/colleges/enrollment_reporting/schedule.php

18 Whitaker, C. (2009). Privacy impact assessment for Central Processing System Washington, DC: U.S. Department of Education. Retrieved from https://www2.ed.gov/notices/pia/cps-032513.pdf

19 U.S. Department of Education. (2016). Common origination and disbursement frequently asked questions. Retrieved from https://cod.ed.gov/cod/downloadCodResource.action?communicationId=96165

20 National Center for Education Statistics. About NPSAS. Retrieved from https://nces.ed.gov/surveys/npsas/about.asp

21 National Center for Education Statistics. About BPS. Retrieved from http://nces.ed.gov/surveys/bps/about.asp

22 National Center for Education Statistics. About B&B. Retrieved from https://nces.ed.gov/surveys/b&b/about.asp

23 Executive Office of the President. (2015). *Using federal data to measure and improve the performance of U.S. institutions of higher education* (p. 24). Retrieved from https://collegescorecard.ed.gov/assets/UsingFederalDataToMeasureAndImprovePerformance.pdf

24 Ibid., p. 27.

25 More than a third of college students transfer. (July 2015). *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/quicktakes/2015/07/08/more-third-college-students-transfer

26 National Center for Education Statistics (2014). Table 303.30: Total fall enrollment in degree-granting postsecondary institutions, by level and control of institution, attendance status, and sex of student: Selected years, 1970 through 2024. *2014 Digest of Education Statistics*, Retrieved from https://nces.ed.gov/programs/digest/d14/tables/dt14_303.30.asp?current=yes

27 Anderson, N. (2014, February 2). U-Md.'s unusual admissions approach: One out of every five freshmen starts in spring term. *Washington Post*. Retrieved from last accessed December 9, 2015, https://www.washingtonpost.com/local/education/u-mds-unusual-admissions-approach-one-out-of-every-five-freshmen-start-in-spring-term/2014/01/31/21d99a64-8794-11e3-916e-e01534b1e132_story.html

28 National Center for Education Statistics. Changes to the 2015-16 IPEDS data collections. Retrieved from https://surveys.nces.ed.gov/ipeds/VisChangesForNextYear.aspx

29 U.S. Department of Education. (2016). Outcome measures: Integrated Postsecondary Education Data System 2016-2017. Retrieved from https://www.regulations.gov/contentStreamer?documentId=ED-2016-ICCD-0020-0002&attachmentNumber=12&disposition=attachment&contentType=pdf

30 IPEDS uses nine racial categories, plus an overall figure, and breaks data down by men, women, and overall.

31 It is not clear how much flexibility FSA has in adding new indicators to NSLDS. The ban on the student-level data system does not appear to lay out restrictions on adding new indicators to students already captured in databases permitted under the law—such as NSLDS. Separately, Section 485B of the Higher Education Act lays out some requirements around NSLDS. This explains the minimum requirements of data in the system but does not appear to preclude the Department of Education (ED) from adding new indicators. Moreover, ED has added new indicators of its choosing to NSLDS in the past. For instance, the gainful employment regulations around career-training programs added indicators for non-federal loans, while the implementation of rules limiting the amount of time students can receive Subsidized Stafford Loans led to requirements for reporting the classification code for students' programs.

32 Author analysis of figures from National Center for Education Statistics (2016). *Integrated Postsecondary Education Data System (IPEDS) 2016-19 and supporting statement part A* (p. 27). Retrieved from https://www.regulations.gov/contentStreamer?documentId=ED-2016-ICCD-0020-0005&attachmentNumber=1&disposition=attachment&contentType=pdf

33 McCann, C., and Laitinen, A. (2014). *College blackout: How the education lobby fought to keep students in the dark* (pp. 7-8). Washington, DC: New America. Retrieved from https://www.newamerica.org/downloads/CollegeBlackoutFINAL.pdf

34 Executive Office of the President. (2015). Using federal data to measure and improve the performance of U.S. institutions of higher education (p. 27). Retrieved from https://collegescorecard.ed.gov/assets/UsingFederalDataToMeasureAndImprovePerformance.pdf

35 Cook, B. (2012). *Incomplete completers: Analysis of a comprehensive graduation rate*. (p. 1). Washington, DC: American Council on Education. Retrieved from https://www.acenet.edu/news-room/Documents/GradRate-110812.pdf

36 Miller, B. (2016). *Scoring the College Scorecard*. Washington, DC: Center for American Progress. Retrieved from https://cdn.americanprogress.org/wp-content/uploads/2016/02/10085654/ScoringScorecard-report.pdf

37 National Student Clearinghouse. (2016). Participating enrollment reporting institutions. Retrieved from http://www.studentclearinghouse.org/colleges/enrollment_reporting/participating_schools.php; National Student Clearinghouse. (2012). Financial aid services brochure. Retrieved from http://studentclearinghouse.info/onestop/wp-content/uploads/FAS_SchoolBrochure.pdf

38 National Student Clearinghouse. (2015). *National Student Clearinghouse enrollment reporting programming and test guide*. (pp. 15-18). Herndon, VA: National Student Clearinghouse. Retrieved from http://www.studentclearinghouse.org/colleges/files/EnrollRept_ProgrammingandTestingGuide.pdf

39 Fishman, R. (2015). *Deciding to go to college*. (p. 8). Washington, DC: New America. Retrieved from http://dev-edcentral.pantheon.io/wp-content/uploads/2015/05/FINAL-College-Decisions-Survey-528.pdf

40 National Center for Education Statistics. (2014). *Profile of undergraduate students: 2011-12*. (p.140). Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubs2015/2015167.pdf

41 Complete College America. (2012). *Remediation: Higher education's bridge to nowhere*. (p. 2). Washington, DC: Complete College America. Retrieved from http://www.completecollege.org/docs/CCA-Remediation-final.pdf

42 Fain, P. (2013, June 5). Remediation if you want it. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2013/06/05/florida-law-gives-students-and-colleges-flexibility-remediation

43 U.S. Department of Education. (2011). *Committee on measures of student success final report*. (p. 17). Retrieved from http://www2.ed.gov/about/bdscomm/list/cmss-committee-report-final.pdf

44 Author analysis of data from the National Postsecondary Student Aid Study 2012 using PowerStats, Table ID cfbbgh2f. Retrieved from https://nces.ed.gov/datalab/powerstats/output.aspx

45 Ibid.

46 Federal Student Aid. (2015). NSLDS enrollment reporting guide. (p. 38). Retrieved from https://www.ifap.ed.gov/nsldsmaterials/attachments/NewNSLDSEnrollmentReportingGuide.pdf

47 Ibid., p. 2.

48 Ibid., p. 318.

49 See note 23.

50 Institute for Higher Education Policy. (2014). *Workforce results data: An excerpt from Mapping the postsecondary data domain: Problems and possibilities technical report*. (p. 3). Washington, DC: Institute for Higher Education Policy. Retrieved from http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/workforce_data_primer_final.pdf.pdf

51 U.S. Department of Education. (2014, October 31). Program integrity: Gainful employment; Final rule. *Federal Register, 79* (211). Retrieved from https://www.federalregister.gov/articles/2014/10/31/2014-25594/program-integrity-gainful-employment#p-907

52 Stone, C., and Chen, W. (2014, June 30). Introduction to unemployment insurance. Center on Budget and Policy Priorities. Retrieved from http://www.cbpp.org/research/introduction-to-unemployment-insurance#_ftnref17

53 Bransberger, P. (2014). *A glimpse beyond state lines*. (p. 17). Boulder, CO: Western Interstate Commission for Higher Education. Retrieved from. http://www.wiche.edu/info/longitudinalDataExchange/publications/MLDE_GlimpseBeyond.pdf

54 These data would not be available if the number of distressed borrowers was too low or too similar to another cohort of students that already had their income information generated.

55 Earnings data would only be available if the number of students who transferred was large enough or different enough from other cohorts that already had earnings information generated.

56 National Center for Education Statistics (2015). Applying for a restricted-use data license. Retrieved from https://nces.ed.gov/statprog/instruct_apply.asp?type=rl

57 National Center for Education Statistics (2015). IPEDS 2015-16 data collection system. Retrieved from https://surveys.nces.ed.gov/ipeds/

58 National Center for Education Statistics. (2015). Admissions. *Integrated Postsecondary Education Data System (IPEDS) 2016-19 and supporting statement part A*, p. 27. Retrieved from https://surveys.nces.ed.gov/IPEDS/Downloads/Forms/package_14_102.pdf

59 *Integrated Postsecondary Education Data System (IPEDS) 2016-17 supporting statement part A*, p. 27.

60 The total represents the sum of hours attached to the six identified components, plus half of the student financial aid survey's burden. Savings for some institutions would be three hours less if they did not fill out the admission survey. Ibid.

61 Perrin, C. (2008, June 30). The CIA triad. *TechRepublic*. Retrieved from http://www.techrepublic.com/blog/it-security/the-cia-triad/.

62 National Institute of Standards and Technology website. (2016). Retrieved from http://www.nist.gov/

63 U.S. Department of Education. (2011). FERPA guidance for students. Retrieved from http://www2.ed.gov/policy/gen/guid/fpco/ferpa/students.html

64 National Student Clearinghouse. (January 2015). Institute for Higher Education Policy. Retrieved from http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/nsc_final_0.pdf.

65 Federal Business Opportunities. (N.D.). Attachment 2: System processing volumes. Retrieved from https://www.fbo.gov/utils/view?id=c6cf50432bec2d60d897c299ea9697ef

66 Fain, P. (2013, March 21). Now what? *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2013/03/21/gainful-employments-future-uncertain-after-court-ruling

67 Transaction details. (2014, March 25). Procurement no. EDFSA09O0056, Modification no. 76, USASpending.gov. Retrieved from https://www.usaspending.gov/transparency/Pages/TransactionDetails.aspx?RecordID=4AF6A036-780C-F191-321E-E13CE91D4914&AwardID=29039393&AwardType=C

68 Transaction details. (2013, September 5). Procurement no. EDFSA09O0056, Modification no. 61, USASpending.gov. Retrieved from https://www.usaspending.gov/transparency/Pages/TransactionDetails.aspx?RecordID=F57284B8-A509-8FFA-8A2D-356F5B131E9B&AwardID=29039393&AwardType=C; Transaction details. (2013, September 11). Procurement no. EDFSA09O0056, Modification no. 63, USASpending.gov. Retrieved from https://www.usaspending.gov/transparency/Pages/TransactionDetails.aspx?RecordID=3A3EA3BC-935E-44FB-1250-E6C907CC22FC&AwardID=29039393&AwardType=C

69 Larger changes appear to cost about $500,000 to $600,000. For instance, in April 2013 ED spent $588,000 to make enhancements to borrower enrollment status, while a year earlier it spent $841,000 to make changes related to implementing gainful employment. Transaction details. (2013, April 5). Procurement no. EDFSA09O0056, Modification no. 50, USASpending.gov. Retrieved from https://www.usaspending.gov/transparency/Pages/TransactionDetails.aspx?RecordID=93960736-23FF-DDE7-F906-3BD44C2F1F3F&AwardID=29039393&AwardType=C; Transaction details. (2012, April 12). Procurement no. EDFSA09O0056, Modification no. 27, USASpending.gov. Retrieved from https://www.usaspending.gov/transparency/Pages/TransactionDetails.aspx?RecordID=68FB63BE-ADF4-D817-E634-8B8500EF8A40&AwardID=29039393&AwardType=C

70 Transaction details. (2012, April 12). Procurement no. EDFSA09O0056, Modification no. 27, USASpending.gov Retrieved from https://www.usaspending.gov/transparency/Pages/TransactionDetails.aspx?RecordID=68FB63BE-ADF4-D817-E634-8B8500EF8A40&AwardID=29039393&AwardType=C

71 See note 6.

72 National Association of Student Financial Aid Administrators. (2015, June 3). Letter from Justin Draeger to James Runcie. Retrieved from http://www.nasfaa.org/uploads/documents/20150603_NASFAA_Letter_Enrollment_Processing.pdf

73 Federal Student Aid. (2015). NSLDS enrollment reporting guide. (p. 67). Retrieved from https://www.ifap.ed.gov/nsldsmaterials/attachments/NewNSLDSEnrollmentReportingGuide.pdf

74 Federal Business Opportunities (2014). Attachment 1: NSLDS statement of objectives—amended. (p. 22) Retrieved from https://www.fbo.gov/utils/view?id=05dfcf7f45701448f3e869ea7da314f5; Konkel, F. (2013, December 12). COBOL still integral to government systems. *Federal Computer Week*. Retrieved from https://fcw.com/articles/2013/12/12/cobol-legacy.aspx

75 Briefcase Systems Development, Inc. (n.d.). *Bloomberg Business*. Retrieved from http://www.bloomberg.com/profiles/companies/9960548Z:US-briefcase-systems-development-inc

76 Federal Business Opportunities. (2014). Phase 1 solicitation: National Student Loan Data System. Retrieved from https://www.fbo.gov/?s=opportunity&mode=form&id=1023f600f0f338da985910a304e6ef9c&tab=core&_cview=0

77 Performance-based organization for delivery of federal student financial assistance. 20 U.S.C. 1018, last amended July 1, 2009. Retrieved from https://www.law.cornell.edu/uscode/text/20/1018

78 Federal Student Aid. (2015). Annual report FY 2015. (p.4). Retrieved from https://studentaid.ed.gov/sa/sites/default/files/FY_2015_FSA_Annual_Report_official.pdf

79 The most recent full year of NSLDS maintenance appears to have cost $11.2 million. Transaction details. (2013, June 27). Procurement no. EDFSA09O0056, Modification no. 57, USASpending.gov. June 27, 2013. Retrieved from https://www.usaspending.gov/transparency/Pages/TransactionDetails.aspx?RecordID=6A7213C7-B7F0-A58F-6B82-FC0351024B95&AwardID=29039393&AwardType=C

80 Ingalls, K. (2006, October 3). Memo to OMB re: Previous terms of clearance. Federal Student Aid. Retrieved from http://www.reginfo.gov/public/do/PRAViewDocument?ref_nbr=200609-1845-003

81 National Center for Education Statistics. (n.d.). NCEES data usage agreement. Retrieved from https://nces.ed.gov/surveys/international/ide/datausageagreement.aspx?link=true

82 National Center for Education Statistics. (n.d.). Technical review panel. Retrieved from https://edsurveys.rti.org/IPEDS_TRP/index.aspx

83 National Center for Education Statistics. (2016). NCES staff. Retrieved from https://nces.ed.gov/staff/stafflist.asp

84 Institute of Education Sciences. (n.d.). About IES: Connecting research, policy and practice. Retrieved from https://ies.ed.gov/aboutus/

85 See note 6.

86 Common Education Data Standards. (n.d.) What is CEDS. Retrieved from https://ceds.ed.gov/whatIsCEDS.aspx

87 Author analysis of data for the 2005-06 academic year. Federal Student Aid. (2006). Title IV program volume reports. Retrieved from https://studentaid.ed.gov/sa/about/data-center/student/title-iv

88 Author analysis of data for the 2014-15 academic year from Federal Student Aid. (2015). Title IV program volume reports. Retrieved from https://studentaid.ed.gov/sa/about/data-center/student/title-iv. Figures for campus-based aid are from the 2013-14 academic year.

IHEP
INSTITUTE FOR HIGHER EDUCATION POLICY