

EXECUTIVE SUMMARY

Building a Student-Level Data System

BEN MILLER

CENTER FOR AMERICAN PROGRESS

MAY 2016

This publication is based on research funded in part by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

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.

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&isposition=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>

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

