Using—and Improving— Federal Student Aid Data Systems to Support Policy Analysis

MATTHEW SOLDNER
AMERICAN INSTITUTES FOR RESEARCH

COLLEEN CAMPBELL
ASSOCIATION OF COMMUNITY COLLEGE TRUSTEES

MAY 2016

POSTSEC DATA

Matthew Soldner, Ph.D., is a principal researcher at American Institutes for Research. **Colleen E. Campbell** is senior policy analyst at the Association of Community College Trustees.

Executive Summary

As the amount of student loans originated each year continues to rise and concerns about loan debt continue to loom large in the minds of the public, policymakers, and advocates, questions about how students finance postsecondary education will continue to rank high among all stakeholders.[1] Management of the U.S. Department of Education's (ED's) federal student aid programs falls to the Office of Federal Student Aid (FSA). In addition to overseeing its trillion-dollar loan portfolio—nearly $980 billion of which is owned and managed by ED—FSA is responsible for administering almost $100 billion in new loans and more than $30 billion in federal grants annually.[2,3]

FSA maintains more than 30 data systems, each in support of the business activities required to successfully operate one of the largest consumer lending and cash transfer operations in the country. Of those 30, four are essential components of the nation's postsecondary data infrastructure. They include the Postsecondary Education Participants System (PEPS), the Central Processing System (CPS), the Common Origination and Disbursement (COD) system, and the National Student Loan Data System (NSLDS). The amount of data housed in these systems is massive; in January 2014, NSLDS alone contained over 30 billion records, and thousands (if not millions) of new transactions occur every day.[4]

These data are a vitally important component of the nation's postsecondary data infrastructure, poised to contribute to our understanding of vexing questions of postsecondary policy and practice:

▶ How can federal student aid programs be best leveraged to promote college access?
▶ How do these programs affect students' persistence and completion outcomes?
▶ After leaving postsecondary education, how do students balance their education debt with other obligations, and how does debt affect their decisions about their personal and professional lives?
▶ Finally, how can aid program administration be strengthened to maximize efficiency, reduce waste, and improve student experience?

Unfortunately, FSA's rich data rarely end up in the hands of the policy analysts and academic researchers who could make effective use of them. Those with the greatest access to FSA data systems are employees of ED, who make "routine use" of the data for the business of government, including the administration of the aid programs themselves, planning for the programs' sustainability, and monitoring the health of the student loan portfolio. Colleges and universities and their partners, such as loan servicers and third-party vendors (e.g., the National Student Clearinghouse), also have varying degrees of access due to their role in the administration of aid programs. Others—including policy analysts in state governments, advocacy organizations, membership organizations, and nonprofit research firms—are often not so lucky.

Recent policy studies and efforts at improving consumer information have demonstrated that when FSA data are put in the hands of qualified researchers, guided by compelling questions, they can be incredibly powerful. This includes helping to better understand the possible causes and correlates of student loan default over time, as well as trends in borrowing and repayment among specific populations of students, such as those attending community colleges.

Systematic solutions meeting a variety of analytic goals are needed if FSA data are to make a larger impact in the policy analysis—and policymaking—community. Those solutions exist; they run the gamut from simply making more policy-relevant tabular data available through existing distribution channels, such as the FSA Data Center, to adopting models, like the Census Research Data Center (RDC), that move data from warehouse to storefront in a secure fashion. Specific solutions include the following:

▶ Develop a systematic process for soliciting feedback about the types of FSA data-based analyses that would benefit policymakers and analysts, and posting the resulting information to the FSA Data Center.
▶ Improve the usefulness of existing NSLDS-based reports provided to campus-based aid administrators, ensuring their contents support local research designed to improve the management of aid programs.
▶ Build the capacity of the nascent FSA Data Office to respond to more complex research requests, and making explicit FSA's commitment to supporting research designed to improve aid administration and policymaking.
▶ Leverage ED's existing *PowerStats* web tool to facilitate secure analysis of new or existing extracts from FSA data systems.
▶ Use ED's existing restricted-use data licensing process to make FSA data extracts or ED policy analysis tools (e.g., Pell Estimation Model) available to qualified researchers.

▶ Explore whether access to ED's new Enterprise Data Warehouse, which brings together data from several key FSA systems, might be possible under the Census Bureau's RDC model.

In a data environment where an increasing amount of information is in the hands of qualified analysts—be it through state longitudinal data systems, federated student unit record systems operating on a regional basis, improved use of National Center for Education Statistics (NCES) sample surveys, or further leveraging proprietary data systems like those held by the National Student Clearinghouse—today's limited access to the information housed in FSA data systems cannot continue. Too many researchers are already arguing to too many stakeholders that *their* critical question could be answered if only they had better data from the Department, and too many important questions about policy and practice have gone unanswered because these data are unavailable. External political pressure will only continue to grow, and may have already become too great to ignore. FSA, well aware of the situation it finds itself in, must act now to find the most effective, efficient, and privacy-protected ways to make information available to those calling for it—and then let them know their voices have been heard.

Using—and Improving—Federal Student Aid Data Systems to Support Policy Analysis

Introduction

The federal government's largest single investment in higher education—and its greatest liability—is a set of federal student aid programs authorized under Title IV of the Higher Education Act of 1965 (HEA), as amended. This includes the well-known Pell Grant program, as well as a series of loan programs including Direct Subsidized and Unsubsidized Loans, as well as Direct PLUS Loans for graduate students and the parents of undergraduates. Management of all of of the U.S. Department of Education's (ED's) federal student aid programs falls to FSA, an operating unit within ED.

Today, FSA is responsible for a student loan portfolio in excess of $1.2 trillion dollars, taken out by nearly 42 million borrowers.[5] Already second only to home mortgages as the largest form of household debt, $97 billion in new federal student loans were disbursed in fiscal year (FY) 2015.[6,7] As a result, nearly one in six adults aged 18 or older has a student loan balance.[8] And many who do are in financial trouble: Approximately 12 percent of outstanding student debt is either delinquent or in default.[9]

In addition to overseeing its trillion-dollar loan program—nearly $980 billion of which is owned and managed by ED—FSA is responsible for almost $32 billion dollars in nonloan programs annually. This includes grant aid for low-income students, most notably $30 billion in Pell Grants, as well as nearly $1 billion in work-study funds distributed to institutions. Before any of those billions are disbursed, however, students must apply for federal assistance. And each year, millions of students—nearly 20 million in FY 2015—do just that by completing FSA's *Free Application for Federal Student Aid (FAFSA)*.[10]

The resulting amount of data is massive. And although just *how* massive is unclear, some simple math gives us a sense of scale: 20 million annual FAFSA records; records detailing 35 million new loan disbursements each year[11]; add Pell Grants to that total, and another 8.5 million recipients.[12] Even before considering data on *existing* student loans, the tally reaches about 64 million records, representing both students and individual loan and grant transactions annually. As of FY 2015, FSA reports its portfolio included 193 million loans. Assuming monthly loan status updates, an additional 2.3 billion records

a year could be added. All told, we estimate FSA data systems ingest nearly 2.5 billion student- or loan-level data records annually. Considering that most loans are in repayment for at least 10 years, we can assume that tens of billions of records are at ED's disposal.

Data collected by FSA are a vitally important component of the nation's postsecondary data infrastructure. Each record maintained by FSA represents a piece of evidence needed to answer some of today's most vexing questions of postsecondary policy and practice:

▶ How can federal student aid programs be best leveraged to promote college access?
▶ Once students have begun their college careers, how does student aid affect their persistence and completion behaviors?
▶ After leaving postsecondary education, how do students balance their education debt with other obligations, and how does debt affect their decisions about their personal and professional lives?
▶ How can institutions and ED strengthen aid program administration to maximize efficiency, reduce waste, and improve student experience?
▶ How can third parties that support the administration of aid programs, such as servicers and debt collection agencies, improve their practices to support loan repayment, thereby reducing rates of delinquency and default?

Unfortunately, FSA's rich data rarely end up in the hands of researchers who could make effective use of them. Questions about ED's annual, multibillion dollar investment in postsecondary education go unanswered largely because of administrative hindrances to data access. Some of these barriers are real, proscribed by law and regulation borne from a legitimate government interest in protecting student privacy. A major constraint is FSA's designation as a performance-based organization (PBO) in the 1998 HEA Amendments, which makes FSA's primary focus the administration and oversight of student aid programs, not necessarily their evaluation.[13] Other barriers are less proscribed, seeming to be as much a function of inevitable confusions about data policy as they are a signal of the priority FSA places on making data available to external researchers.

There are solutions—some small, some large—to this problem. To understand those solutions, and to make the value proposition for their adoption, we undertake five activities. Initially, we briefly describe four key FSA data systems that are likely to be of greatest value to researchers. Then, we discuss the legal and regulatory framework that manages access to FSA data resources. Next, we highlight ways governmental and nongovernmental researchers have leveraged these data to answer important questions of policy and practice. We conclude by offering a set of action items that, if enacted, can unlock the promise these data represent.

FSA Data in a Nutshell

FSA maintains more than 30 data systems, each in support of the business activities required to successfully operate one of the largest consumer lending (and cash transfer) operations in the country.[14] Of those 30, four are essential components of the nation's postsecondary data infrastructure. They are (1) PEPS, (2) CPS, (3) COD, and (4) NSLDS. We briefly describe the contents of each system in **Table 1**. A link to the most recent versions of each system's technical reference is included in **Table 2**.

PEPS, the Postsecondary Education Participants System
As its name implies, PEPS is FSA's primary repository for storing institution-level data on entities involved in administering federal student aid, including colleges and universities, and student loan lenders, guarantors, and servicers. The backbone of PEPS is the School Profile, which includes basic demographic information for institutions, each represented by a unique six-digit Office of Postsecondary Education Identification number (OPEID). This includes information about the campus' main and branch locations (if any, represented by two-digit suffixes to the six-digit OPEID); type (e.g., public) and longest program length; current participation in various federal aid programs; accreditation and state authorization status; ownership status for proprietary institutions (e.g., publicly held corporation); key contacts; and recent default rates. A subset of School Profile data are available for download from FSA.[15]

There is more to PEPS, though, than its downloadable dataset would suggest. Of particular interest are data elements related to institutions' financial health. Some of these are available elsewhere on ED's website, such as financial responsibility composite scores and heightened cash-monitoring status. Others are not routinely available, including the accounting data that drive ED's financial responsibility metrics, as well as records detailing program and audit reviews, such as findings of deficiencies, required repayments, and management improvement plans. These data could be provided to consumers in order to provide a more complete picture of the institution's health and stability. ED has signaled the importance of some of these data by including a flag for institutions that are on a heightened cash-monitoring status on the new College Scorecard. PEPS also contains lender and guarantor audit information and default rates.[16]

CPS, or Central Processing System
CPS verifies information provided by students on their FAFSA and performs the federal needs analysis used by financial aid administrators to determine a student's eligibility for federal aid programs.[17] CPS queries several other federal data systems to ensure student eligibility for aid programs. Computer matches include identity verification with the Social Security Administration, male applicants' registration with the Selective Service, noncitizens' registration with the Department of Homeland Security, and veteran's status with the Department of Veterans Affairs. CPS checks against other FSA systems to guard against grant overpayment and the disbursement of new loans to students who already have education loans in bankruptcy. The results of these matches, including to citizenship status, Selective Service registration, NSLDS history, and veteran's status, become part of a student's CPS record.

Assuming an applicant's FAFSA passes these screening processes, CPS uses the asset and income information provided by the student to automatically calculate a student's Expected Family Contribution (EFC). The result of processing, a student's Institutional Student Information Record (ISIR), is then transmitted electronically to the school or schools of the student's choice so that financial aid administrators can insert it into their financial aid and student information system and package the student's financial aid. CPS stores the first application a student submits in a given award year, referred to as the "01 transaction." Edits to that transaction are stored in CPS and numbered sequentially (e.g., a "02 transaction").

As a result, an aid applicant's records in CPS contain detailed demographic and financial data. This includes an applicant's full name and Social Security number; contact information (including e-mail address); citizenship status; sex; marital status; parental educational attainment; high school; intended award level (e.g., certificate, associate's degree, bachelor's degree); income, asset, and tax data for the student, and, if applicable, parents and spouse; military status; status as an orphan or ward of the court; homelessness sta-

tus; and use of social service benefits (e.g., Supplemental Nutrition Assistance Program, Free or Reduced Price Lunch, Temporary Assistance for Needy Families, or Women, Infants, and Children). It also includes values for each variable used in the student's needs analysis, culminating in a student's EFC.[18] CPS data are necessarily personally identified, and they are considered highly sensitive by ED.

Despite being self-reported, the quality of key CPS data elements is generally high. The computer matching protocol identified above provides a strong check against many errors, and the IRS Data Retrieval Tool (when used) helps to ensure that the tax return data requested on the FAFSA are entered without error. To help ensure accuracy of CPS data, a proportion of FAFSAs are selected for verification each year. Institutions are also responsible for verifying FAFSA data they "have reason to believe may be incorrect."[19] Generally, verification involves visually inspecting tax form transcripts, W-2s, or tax returns themselves to ensure what students have entered on their FAFSA is correct. Other items that may be identified for verification are the number of household members who are in college, high school completion, and use of food stamp benefits.

TABLE 1. DATA ELEMENTS COMMONLY FOUND IN KEY FSA DATA SYSTEMS

Data Element	PEPS	CPS	COD	NSLDS
Institutional characteristics	✔			
Institutional financial health	✔			
Institutional audits and program reviews	✔			
Lender and guarantor performance information	✔			✔
Student identifying information		✔	✔	✔
Student and parental academic attainment		✔		
High school		✔		
Award level for current program of study		✔	✔	✔
Financial and benefit program participation data needed for federal needs analysis		✔		
Dependency status for federal student aid		✔	✔	
EFC		✔		
Enrollment status and intensity			✔	✔
Grant disbursements and amounts			✔	✔
Loan disbursements and amounts			✔	✔
Field of study			✔	✔
Grant and loan eligibility used			✔	✔
Loan repayment status detail, including default and delinquency				✔
Loan repayment plan				✔
Loan terms detail				✔
Loan servicing data				✔
Consolidation/consolidated loan status				✔

NOTES: For illustrative purposes only. Contents of systems are regularly enhanced. For most recent system documentation, consult source materials listed in Table 2. Not all data elements are available for all loan types; more data are available for loans originated under the Direct Loan program, and more data are reported for subsidized loans than unsubsidized loans (e.g., field of study.)

SOURCE: See Table 2 in this document.

COD, or Common Origination and Disbursement System
After preparing a student's federal student aid package, financial aid administrators at colleges and universities (or their designees, such as third-party servicers) transmit a record to FSA via COD to award, and later disburse, grant and loan aid.[20] COD is also used to update and cancel disbursements, discharge loans, initiate Direct PLUS Loan credit checks and update the disposition of those checks and any related appeals, and perform calculations and notifications related to the 150 percent Subsidized Usage limit for Direct Loan Borrowers.[21]

For all types of federal aid, COD processes basic information about the recipient submitted by the institution, including identifying information for both the student and the awarding institution. Much of these data are automatically populated with FAFSA data by the institution's student information system before being electronically transmitted to COD. Student data include name, Social Security number, date of birth, citizenship status, postal and electronic mail address, and drivers' license information. Pell Grant–specific data include dates of enrollment within a given payment period, whether a student is incarcerated, the student's actual cost of attendance and EFC, the student's enrollment intensity, disbursement dates and amounts, and the proportion of a student's Pell Lifetime Eligibility used. Similar data are maintained when awarding a student loan, including whether the loan is subsidized, unsubsidized, or part of the PLUS program; the student's contact information; additional demographic data including financial aid dependency status and academic class level; enrollment intensity; the six-digit Classification of Instructional Programs (CIP) code and credential level that corresponds to the student's field of study; program length (for Subsidized Loans); and details about the loan's authorized amount, payment period, and disbursement dates.[22] Much like CPS, COD data are personally identified and contain sensitive information.[23]

NSLDS, or the National Student Loan Data System
NSLDS is FSA's primary mechanism for tracking the status of federal student loans and grants, from disbursement to repayment. It also includes data about loan and grant recipients' ongoing enrollment status, including whether they have earned an educational credential. Finally, ED uses NSLDS as a portal for annual gainful employment reporting. Gigabytes of data stream in to NSLDS from institutions, loan servicers and guarantors, and other FSA systems on a daily basis. As a result, it is by far the largest—and the most complex—system in the FSA data architecture.

Aid recipient enrollment reporting occurs at least every two months, and it includes information about a student's enrollment intensity on a given date; the institution's anticipated completion date for a student; and each program in which a student is currently enrolled, including six-digit CIP code, award level, program length, and the date on which the program began. Both to trigger repayment and to calculate students' remaining eligibility for Subsidized Direct Loans, institutions are also required to report when students have withdrawn from or completed a program of study.[24] Historically, the quality of data related to student awards has been inconsistent. Acknowledging the problem, FSA sent a "Dear Colleague Letter" to institutions in 2012, reminding them of their responsibility to ensure these data are both timely and accurate.[25]

Finally, institutions that offer programs subject to ED's gainful employment regulations—which include all programs at for-profit institutions as well as certificate programs at public and private, nonprofit institutions—use NSLDS to comply with annual reporting requirements. For each student enrolled in a program subject to the regulation, institutions must report information about the program of study (e.g., name, six-digit CIP code, length, and award level); dates of enrollment and current enrollment status (i.e., enrolled, withdrawn, or graduated); and, for students no longer enrolled, their date of withdrawal or graduation, the amount of institutional loans and private loans (if known to the institution) the student received during their course of study, the total amount of tuition and fees paid, and the typical cost of books, supplies, and equipment.[26]

The bulk of loan-level detail in NSLDS is synthesized for institutions via preformatted reports that can be generated on an ad hoc basis or scheduled for regular download. Among the more commonly used reports is the School Portfolio Report (SPR). Using their SPR, campus users (or authorized designees)—typically financial aid officers—can access information about all loans associated with their institution, including consolidation loans.

For each loan found on the SPR, an institution's financial aid administrator can access the borrower's name and Social Security Number; the student's academic class level and anticipated completion date at the time the loan was made; the loan date, type, interest rate, and original loan amount; the outstanding principal balance currently owed, in addition to accrued interest and fees; the current loan status and the date that status was reported; the original and current guaranty agency or servicer; the date the loan is expected to enter repayment; deferment dates and status; for federally serviced loans, the most recent scheduled payment amount and payment date by the borrower, their repayment plan and length, and, if applicable, when the loan entered forbearance or delinquency; if applicable, the date the loan entered default; whether the loan was eventually rehabilitated; and loan consolidation information.[27] Loans originated by institutions *not* associated with the user, including consolidated loans originated at other institutions, are not displayed.

TABLE 2. RESOURCE DOCUMENTS FOR KEY FEDERAL STUDENT AID SYSTEMS

System	Most Recent Reference Document	URL
PEPS	PEPS Data Extracts	http://www2.ed.gov/offices/OSFAP/PEPS/dataextracts.html
CPS	ISIR Guide, 2016–2017	https://www.fsadownload.ed.gov/ISIRGuide1617.htm
COD	COD Technical Reference, 2016–2017	https://www.fsadownload.ed.gov/CODTechRef1617.htm
NSLDS	NSLDS Record Layout Library	https://ifap.ed.gov/ifap/byNSLDSType.jsp?type=NSLDS%20Record%20Layouts

How Access to FSA Data Is Managed

FSA data represent a treasure trove to policy analysts with an interest in federal student aid programs as well as federal and state policymakers working to design effective policies, whether their questions center on understanding returns to the federal investment in loans and grants, improving the management of aid programs, or leveraging data to power consumer information tools or drive accountability efforts. But given FSA data's sensitive and, at least in their raw form, personally identified nature, access to them is tightly controlled. These privacy protections have their basis in a series of federal laws, including the Privacy Act of 1974 and HEA (as amended). More information about privacy laws, regulations, guidance, and best practices can be found in Joanna Grama's paper *Postsecondary Education Data Systems: Information Security and Privacy Best Practices*, also in this series.

The Privacy Act of 1974 (Public Law 93–579) requires the federal government to disclose what personally identifiable information it collects, how those data are used, and the conditions under which they may be shared with others. These disclosures, published in the *Federal Register* and cataloged on ED's website, are known as system of records notices (SORNs).[28] By default, the Privacy Act limits the disclosure of personally identifiable data without the written permission of the individual who is the subject of the record, unless one of several exceptions are met, including exceptions for "routine use" and "statistical research."[29]

For some of its systems, FSA has included "research disclosure" as a routine use. The current SORNs for PEPS,[30] COD,[31] and CPS[32] include the following language: "The Department may disclose records from this system of records to a researcher if the Department determines that the individual or organization to which the disclosure would be made is qualified to carry out specific research related to functions or purposes of this system of records. The official may disclose records from this system of records to that researcher solely for the purpose of carrying out that research related to the functions or purposes of this system of records. The researcher must be required to maintain Privacy Act safeguards with respect to the disclosed records."

Research disclosure, at least as a routine use of the data, is notably absent from the SORN for NSLDS.[33] However, the SORN does indicate that a purpose of NSLDS is to "support research studies and policy development."[34] It also notes ED "may disclose records to Federal, State, and local agencies" to "support governmental researchers and policy analysts."[35] Absent this, the Privacy Act's statistical research exception may permit the same type of disclosure as appears to be allowed by the SORNs for PEPS, COD, and CPS. This exception permits disclosure to "a recipient who has provided the agency with advance adequate written assurance that the record will be used solely as a statistical research or reporting record, and the record is to be transferred in a form that is not individually identifiable."[36] Exceptions notwithstanding, it is important to remember that the Privacy Act focuses on when ED *may* release a record, not when it *shall* or *must*.

The Higher Education Opportunity Act of 2008 (Public Law 110–315), notorious for explicitly banning ED's development of a student unit record system, also mandates a particular privacy protection for NSLDS data. In Section 489, the law requires the Secretary to "[prohibit] nongovernmental researchers and policy analysts from accessing personally identifiable information" held in the system. This is, of course, already mandated by the Privacy Act. Importantly, the Family Educational Rights and Privacy Act of 1974 (FERPA) does not bear on ED's capacity to disclose data from FSA systems.[37] FERPA requires educational institutions, not the federal government, to maintain the privacy of educational records.[38]

Who Gets Access to FSA Data, and for What Purpose(s)?

Only one class of individuals has unqualified access to data held by FSA: students, who, per the Privacy Act, may access their *own* NSLDS records via a web-based portal hosted by ED for verifying data accuracy. Otherwise, those with the greatest access to FSA data systems are ED employees. Colleges and universities and their partners, such as third-party servicers, also have varying degrees of access to PEPS, CPS, COD, NSLDS, and other systems due to their role in the administration of aid programs, from disbursement to servicing and collection. Others, including policy analysts in state government, advocacy organizations, and nonprofit research firms, are often not so lucky.

Notwithstanding their use in the actual operation of aid programs, FSA data systems are rarely used in their raw form by governmental researchers. The sheer volume of data, paired with the challenge of querying production systems that are processing hundreds of thousands of transactions per day, means that most analysts work with extracts from FSA systems and that policymakers receive information in tabular form. Often, these analysis products are ad hoc in nature, supporting the development of policy and regulation within the Executive Branch. NSLDS data are known to have been used in the development of regulatory impact analyses associated with gainful employment, for example, and it is widely held that ED has used NSLDS to create—but not release—institution-level default rates for PLUS loans. Some extracts, though, form the basis of data tools that, although critical to analyses that have been conducted by ED and its partners for years, are not routinely publicly available.

The first and most well-known example of this practice is Cost Estimation and Analysis Division's Statistical Abstract (CEAD-STAB). Maintained by ED's Budget Service, CEAD-STAB is a random sample of more than 2 million borrowers from NSLDS, along with their associated loan records. Fully de-identified, these data allow the Budget Service to identify a series of assumptions that inform cost estimates of loan programs, supporting federal budgeting processes both within ED and at the Office of Management and Budget. A second, similar extract exists for CPS. Containing a de-identified sample of nearly half a million FAFSA filers, the Aid Applicant Sample File allows Executive and Congressional Branch analysts to simulate the impact of changes in need-analysis methodologies.[39]

A third resource is the Pell Grant Estimation Model. Also maintained by ED's Budget Service, the Pell Model is driven by data from a nationally representative sample of Pell Grant recipients and aid applicants, and is used to forecast program costs under a variety of applicant, economic, and policy scenarios. Although the Pell Model was once available to qualified researchers, including those working with national higher education associations, no recent versions of the tool are known to exist outside ED, the Congressional Research Service, the Congressional Budget Office, and the Office of Management and Budget.[40]

Few other examples of record-level access to FSA data systems for analytic purposes are known to exist at the federal level. A notable exception is the use of CPS, COD, and NSLDS data to support sample surveys developed by ED's NCES, including the National Postsecondary Student Aid Study, the Beginning Postsecondary Students Longitudinal Study, and the Baccalaureate and Beyond Longitudinal Study. In these cases, FSA data of sample members are appended to each study's analysis file and then fully de-identified. However,

because these studies are typically only nationally representative in nature and are fielded on a relatively infrequent basis (once every four and eight years, respectively), they often fail to answer more detailed—or more time-sensitive—policy questions.

Case Studies of Data Access and What They Teach Us

As two recent examples demonstrate, when FSA data are put in the hands of qualified researchers, guided by compelling questions, they can be incredibly powerful. Doing so also makes more widely transparent the challenges of using FSA data for policy analysis. The first case study uses NSLDS data provided by a set of cooperating institutions; the second, a data source not unlike CEAD-STAB (perhaps CEAD-STAB itself). Both studies demonstrate how research based in FSA data can be leveraged to generate insights about policy and practice at the federal, state, and institutional levels.

A Closer Look at the Trillion: Borrowing, Repayment, and Default at Iowa's Community Colleges

This report, authored by Colleen Campbell (Association of Community College Trustees [ACCT]) and Nicholas Hillman (University of Wisconsin–Madison), and published in September 2015 by ACCT, used data from all 16 of Iowa's community colleges to examine trends in federal student loan borrowing and repayment.[41] The project was spurred by interest from the trustees of Iowa's community colleges, who were concerned about their institutions' persistently high default rates. The institutions asked ACCT to conduct an analysis of their NSLDS and institutional data to highlight areas for targeted interventions.

Each college shared two data files from NSLDS: SPR, mentioned above, as well as the Loan Record Detail Report (LRDR), which is pushed to schools annually and contains information on borrowers included in the college's Cohort Default Rate. Institutions appended data on student characteristics that were either not available from NSLDS or not included on the LRDR or the SPR, such as the student's credits earned; credentials completed; dependency status; estimated family contribution; family income; housing situation; program of study (i.e., CIP code); and satisfactory academic progress.

Despite institutions' relatively unfettered access to NSLDS, some data limitations remained. Most importantly, institutional reports from NSLDS contain data only for loans associated with the school querying the report. If a student borrowed from another institution before (or after) enrolling in one of Iowa's community colleges, those loans do not appear on the institution's report. This presents challenges for researchers interested in student pathways through postsecondary education and total indebtedness of borrowers. Missing data are also a concern, mostly due to different reporting standards for Federal Family Education Loan (FFEL) and Direct Loan ser-

vicers. For example, FFEL servicers are not required to report certain pieces of information, such as dates of delinquency and deferment, loan repayment plans, and last payment date. As FFEL loans are retired, the impact of FFEL-related data quality issues will lessen.

The process by which *A Closer Look at the Trillion* was undertaken highlights a potentially useful, if inconvenient, route to accessing NSLDS data. While ACCT would not have been in a strong position to access NSLDS data on its own—and it would have had no access to the supplemental data provided by institutions—ACCT serving as those institutions' research partner, signing nondisclosure agreements with those colleges, and receiving only de-identified data mitigated Privacy Act concerns. The report's analyses also demonstrate that FSA data can be leveraged to arrive at important insights, especially when combined with institutional data (the same would be true were data provided from a state's longitudinal data system). The authors confirmed prior research on default trends and brought to light more information on loan servicing, an area that received little study until the publication of this report. However, an endeavor such as this one can be a significant undertaking for institutions, which may not have the financial resources and research capacity to study these data. And because the data used only represent Iowa's community college students who entered repayment in FY 2011, it highlights the tension between locally relevant results and readily generalizable insights.

A Crisis in Student Loans? How Changes in the Characteristics of Borrowers and in the Institutions They Attended Contributed to Rising Loan Defaults

This report, authored by Adam Looney (Department of Treasury) and Constantine Yannelis (Stanford University), marks the first time external researchers were able to access a significant subset of NSLDS data directly from ED.[42] The authors write they used a "random 4 percent sample of all federal student loan borrowers," a file that may well have been ED's CEAD-STAB, matched to Department of the Treasury data on borrower's earnings and income.

The analysis of FSA data via something like CEAD-STAB offers several benefits. First, the data are longitudinal, following a cohort of borrowers from when they first took out a federal student loan. The dataset contains complete NSLDS information on borrowers for almost 20 years, allowing the researchers to identify trends over time and providing a significant period postenrollment to study earnings. Second, administrative data derived from a single federal source are virtually by definition less error prone, more complete, and more tractable than would be similar data collected from multiple institutions or systems. Finally, because it is representative of all borrowers, a dataset like CEAD-STAB offers greater coverage of diverse types of students and institutions

than could be had from any other data resource, save a (currently illegal) national student unit record system.

Although the data used by Looney and Yannelis are the most comprehensive to date, they still suffer from limitations. First, although institutions have long been required to report persistence outcomes such as withdrawal or completion to NSLDS, these data were not always reported accurately before 2012. Until recent changes to the length of time students were eligible for unsubsidized loans made tracking these outcomes important to financial aid administrators, who are keen to safeguard students' access to low-cost capital, these data elements were used only to trigger loan repayment, which occurs regardless of whether a student graduated or withdrew. Second, data like that found in CEAD-STAB includes only data about student loan borrowers. While the outcomes of these students are important, CEAD-STAB–like datasets do not offer the research community a similar student-level view of nonaided students, which makes comparisons between aided and nonaided students impossible.

Despite these shortcomings, Looney and Yannelis' work brought additional nuance to important sector-wide variation in students' borrowing and repayment behavior. For example, the authors disentangle a number of factors associated with attendance at for-profit institutions, higher debt, and poor repayment outcomes—the latter an issue prior research had difficulty quantifying. Additional data such as these, or more robust data provided by a linked FSA statistical abstract or federal student unit record system, would provide additional findings that could help decision makers craft more equitable, efficient policies for all postsecondary students.

The Unique Case of the 2015 College Scorecard

Like Looney and Yannelis, ED is also in a position to create hybrid data products that combine information from a variety of systems. With its release of the 2015 College Scorecard, it did just that, combining data from NSLDS, CPS, the Department of the Treasury, and NCES' Integrated Postsecondary Education Data System (IPEDS). The resulting dataset includes nearly two decades of information regarding colleges and universities, their students' federal borrowing behavior, and borrowers' eventual wage outcomes and repayment experiences.

The Scorecard's explicit purpose is to meet the consumer information needs of students and families, helping them "compare colleges and see how well schools are preparing their students to be successful."[43] To that end, it provides consumers—and consumer-information tool developers—data never before released by ED. This includes (for various cohorts of students) the median federal debt of federally aided exiters, disaggregated by completion status; NSLDS-derived transfer and completion rates at two, three,

four, and six years; mean and median earnings for students six and 10 years after their initial enrollment; the proportion of students meeting a $25,000 annual earnings threshold; and student loan repayment rates one, three, five, and seven years after entering repayment status. Many of these measures are available at policy-relevant levels of disaggregation, including income band, Pell-recipient status, and status as a first-generation college student.

Although a significant advance in consumer information, critics often note that the College Scorecard is beset with a significant flaw: Much of the new information it provides pertains only to students who receive federal aid. While this is true, a majority of today's students are aided, up to 70 percent nationally. This inclusion of only aided students is problematic for some measures, like earnings, but is not a limitation for debt-related measures (e.g., debt, repayment) that apply only to borrowers. The Scorecard also focuses on the outcomes of undergraduate students, which, for institutions that are largely (or wholly) graduate-serving, limits its utility. Also restricting the utility of the dataset is the large number of null or privacy-suppressed values present, which make certain data elements difficult to analyze, especially for smaller institutions and colleges where few students receive aid. Finally, some of the Scorecard's most interesting data—such as wage outcomes—are available only at the institution level. Prior research, including work done by American Institutes for Research, has demonstrated that wage outcomes can be as variable within an institution, between students in different academic programs, as they are across institutions.[44]

Action Items for Improving Access to FSA Data

As the examples above evidence, it *is* possible to access FSA data for research, accountability, and consumer information purposes. But it is far from simple: Not every qualified analyst with a compelling research question will have the capacity to rally dozens of institutions or leverage their role as a government analyst to access not one but *two* otherwise highly restricted sets of data. And even when they can, compelling policy questions go unanswered. Systematic solutions that meet a variety of analytic goals, while maintaining appropriate privacy protections, are needed if FSA data are to become more consequential in the policy-analysis and policymaking communities.

Luckily, there are indications that FSA has begun to place an increased emphasis on putting information from its data systems—if not the data itself—in the hands of policy analysts. As we discuss below, FSA is in the process of developing a data office, reporting to the chief operating officer, to increase its capacity to respond to ad hoc information requests. There are also nascent plans to leverage a new project, the Enterprise Data Warehouse and Analysis initiative, to create new opportunities for conducting policy-focused research. Below,

we discuss these and other solutions that vary in complexity, cost, and analytic utility, including specific action items that begin to make those solutions a reality.

Improving the FSA Data Center

Most, if not all, aid policy analysts are familiar with the FSA Data Center, a web portal through which ED shares static, tabular data on federal student aid programs. Among the most used reports in the Data Center are quarterly summaries of institutions' loan and grant-making activity, including the number of students served and total dollars disbursed by program (e.g., unsubsidized Direct Loans, subsidized Direct Loans, and PLUS Loans for Graduate Students). Quarterly reports that detail the status of ED's loan portfolio are a welcome new addition. Each dataset allows researchers to analyze trends in aid use nationally, by state or institution sector, and at individual institutions.

Although the Data Center is an invaluable resource, there is room for improvement. We have four concerns. First, because the unit of analysis is a grant or a loan, rather than a student, it is not possible to make any statement about an "average student's" financial aid package. Second, within a loan program, most data are aggregated at the level of the institution, making it impossible to understand subpopulations that might be of particular policy interest, such as lower-income students or students from various demographic backgrounds. Third, the aggregation of undergraduate and graduate students in program totals is particularly problematic, making student-level trends in borrowing difficult to discern and the effects of differences in program pricing impossible to disentangle. Finally, depending on the specific report one is reviewing in the Data Center, data can be presented for the entire Title IV portfolio, by sector, or by institution, which makes synthesizing information across the reports nearly impossible. Presumably, each of these shortcomings is relatively easily overcome in FSA data systems. It just isn't done—or, if it is, it isn't made public.

Action item

A simple, incremental solution to improving the usefulness of FSA data is expanding the contents of the FSA Data Center. After all, there is nothing wrong with a static table if the data that table contains have information that answers a compelling policy question or is useful to an individual institution. We envision two, non–mutually exclusive routes to making this recommendation a reality.

One option for FSA to consider is the development of a process to regularly solicit input from the policy research community about the types of analyses and information that would add the greatest value to the community's work. An annual request for information (RFI) would be both quick and cost effective; it has the added benefit of being predict-

able. In conjunction with a technical review panel of national experts, FSA could adjudicate and prioritize requests that arise from the RFI, and then task FSA staff to generate and post the results to the FSA Data Center. If these requests yielded analytic variables at the institution level that could be crosswalked to IPEDS institutional identifiers, made easier by an OPEID-UNITID crosswalk released as part of the College Scorecard project, the relevant data could also be added to the IPEDS Data Center or the Scorecard for wider use.

A second option is, conveniently, an approach that senior FSA officials have indicated is already in the offing: using FSA subject matter experts and analysts assigned to the newly constituted FSA Data Office to respond to a broader and more complex array of ad hoc research requests. Although many details remain unresolved, the potential of this solution to bolster the relationship between FSA and the policy research community is undeniable. The creation of a cadre of FSA experts who have the explicit responsibility to respond to researcher information requests sends a powerful message about the importance FSA places on leveraging its data for policy analysis.

To be sure, none of these solutions are free: Convening experts and paying staff takes resources. But these costs are likely to be incremental at best—unlikely to cost hundreds of thousands of dollars a year. Meetings can be held virtually, and new, well-qualified analysts (if needed) can be hired at all levels along the federal government's salary schedule. Given the many demands FSA faces each day, the resource most critical to implementing this solution is its continued prioritization by senior ED leadership. Advocates and analysts who believe this is a potentially useful mechanism for using FSA data to improve policymaking must demand it.

Improving Information for Institution-Based Researchers
Financial aid administrators already have access to a wide range of preformatted reports, largely based on data from NSLDS. This includes (1) the SPR, mentioned above, which focuses on the current status of individual loans; (2) the LRDR, which focuses on loans entering repayment as part of a specific repayment cohort; and (3) the Delinquent Borrower Report, which details borrowers whom servicers have reported as delinquent in their repayment. Each contain data that can be useful resources to aid administrators as they seek to effectively manage borrowing on campus, and support students entering loan repayment. They can also support data-driven decision making if they contain the information of most value to campus research professionals who seek to answer questions related to improved aid administration.

Action item
FSA should extend its collaboration with aid-related professional associations, institution-based aid researchers, and other stakeholders to ensure that its current preformatted reports support research designed to improve campus leaders' ability to effectively manage their student aid dollars. We do not presume to know what questions are most pressing to the community aid research practitioners as a whole, but anecdotal reports from individual professionals suggest they run the gamut from using predictive analytics in delinquency and default avoidance to better understanding how their students' borrowing behaviors compare with those at peer institutions. Luckily, there is no need to guess what sort of information would best support this effort: FSA can rely on a well-used ED approach, the Technical Working Group (TWG), to find out. Annual TWGs specifically focused on ensuring that the data FSA returns to institutions participating in Title IV programs is useful and actionable, perhaps as part of existing FSA-sponsored or related conferences, would be a simple and low-cost way to gather that feedback.

Leveraging Extracts From FSA Data Systems
Static tables are unlikely to answer every question posed by aid researchers and policymakers. Indeed, much of the government's own work—and work like that of Looney and Yannelis—demonstrates that analysts do not always need direct access to FSA systems in order to perform important research. Thoughtfully developed and fully de-identified extracts, like CEAD-STAB, can answer a range of policy questions. Extracts can be leveraged in two non–mutually exclusive ways.

Action item
First, FSA could work with NCES to use its *PowerStats* tool to permit secure tabular and correlational analysis of a de-identified extract like CEAD-STAB. *PowerStats* is already used thousands of times a year by the general public to run analyses on NCES datasets, such as the restricted-use data files associated with the National Postsecondary Student Aid Study (NPSAS) family of studies, vastly expanding opportunities for the analysis of that data without running any risk of disclosing personally identifiable information. It is important to note that the implementation of a *PowerStats for FSA* solution would be both more complex and more costly than simply putting additional FSA-generated tables on the FSA Data Center.

The first challenge is that current extracts contain many more records than do NCES restricted-use data files, often by a factor of 10 or more. This would demand high-capacity, high-speed systems and the optimization of the underlying *PowerStats* code to leverage them. However, because *PowerStats* is currently provisioned by resources from Amazon's GovCloud (a cloud hosting solution offered by Amazon Web Services that complies with a series of federal requirements governing information assurance), appropriate systems are available virtually on demand.[45,46] Equally as complicating is the need for extracts

placed within the *PowerStats* environment to be formatted to comport with its existing analysis architecture. *PowerStats* expects a single record for each student; however, given the transactional nature of most FSA data systems, a single student could be linked to dozens of NSLDS and CPS records. Reengineering existing extracts is possible, but it would require time, expertise, and a significant initial monetary investment. If the cost of this activity were relatively equivalent to what NCES spends annually to improve and maintain its current *PowerStats* suite of tools, then we might expect the cost to approach $1 million annually.[47]

Action item

An alternative—and vastly simpler—solution would be to license extracts to qualified researchers via a system not unlike NCES's restricted-use data licensing process. NCES has already established a rigorous application, screening, and security-monitoring process for safely sharing dozens of de-identified datasets with hundreds of government, academic, and nonprofit researchers across the country. After a thorough review for potential disclosure risks, CEAD-STAB, the Applicant Sample File, the Pell Model, or any other extract of FSA's choosing could simply be added to the menu of data available from ED, to be accompanied by detailed data documentation. Because this solution does not require substantial manipulation of existing data resources, it could be implemented at little cost to FSA.

Both the *PowerStats* and restricted-use solutions offered here place a significant amount of FSA data in the public sphere. This is beyond FSA's current remit, which is primarily concerned with delivering federal student aid programs in a manner that is free of waste, fraud, and abuse. By facilitating the analysis of data that has historically been the sole property of the government, ED also opens itself to new questions about the conduct of federal aid programs.

Given the current tenor of conversations surrounding FSA, ED may be disinclined to move forward with any solution that places analysis in the hands of nongovernmental policy researchers, unless compelled to by Congress. A recent Government Accountability Office report found that FSA's guidance to loan servicers "are sometimes lacking, resulting in inconsistent and inefficient services to borrowers" and that "[w]ithout improved guidance and instructions to servicers, borrower finances or the integrity of the Direct Loan program could be negatively affected."[48] Members of Congress have also criticized FSA's management practices, including a recent statement by Rep. Virginia Foxx before the House Committee on Education and the Workforce, in which she characterized FSA as "rife with inefficiencies that have led to ...inaccurate reporting of data."[49]

Providing Access to the FSA Enterprise Data Warehouse

The most sweeping route to increasing access to FSA data would be to provide researchers a mechanism to query the system directly. Unfortunately, this is not feasible for several reasons. First, there are clear Privacy Act prohibitions against ED disclosing the personally identifiable data contained in FSA systems. Second, performing queries while systems are "live" degrades ED's capacity to conduct routine business operations. Finally, because FSA data systems are transactional, there is never a singular, unchanging FSA dataset, making validation and replication of prior analyses difficult if not impossible.

A potential solution comes with the advent of ED's anticipated Enterprise Data Warehouse and Analytics project. Designed to facilitate analysis by governmental researchers both inside and outside ED, the Warehouse is meant to provide a single point of access to a (comparatively) static set of data drawn from PEPS, CPS, COD, NSLDS, and other critical business systems. Developed by Accenture and already in limited use by FSA, the Warehouse is meant to provide government analysts "timely, accurate, consistent, and repeatable access to FSA data."[50] According to the president's 2017 budget, the cost of that access is $12.3 million this year, with $5.8 million allocated for system operations and maintenance and $6.5 million for development work.[51] Most importantly, however, is the assertion in the FY 2016 budget that those monies will be used to "make student aid data more accessible for large scale research by external stakeholders."[52]

Action items

The pivotal question surrounding the Warehouse is whether it will make *data* available, or whether what is made available is *information*. The former implies researchers would have relatively direct access to de-identified student-level data and the analytic tools to conduct their own analyses; the latter that policy analysts would pose research questions to FSA subject matter experts and receive "results" in return. Either route offers important benefits to the policy research community.

Providing relatively direct researcher access to the data warehouse would not be unprecedented. ED should query other federal agencies for how they have developed physical (or virtual) data centers that facilitate secure research and analysis activities. Although several models exist within government that could serve as inspiration for ED's next steps in data policymaking, the Census Bureau's RDC is among the best known.

The premise of an RDC is simple: After agreeing to a series of security requirements, researchers with approved research plans travel to a physical data center where they conduct their analyses. As needed, researchers can augment data stored in an RDC with other administrative data to which they

have access. All analyses are conducted within the RDC's virtual environment using commonly available software tools. To ensure the security of the data housed in the RDC, only prescreened, tabular output is permitted to leave. The cost of operating the RDC program cannot be discerned from the Census Bureau's most recent budget request.[53] Were ED to begin a similar activity with a single site (presumably Washington, DC), personnel and infrastructure costs could easily reach the $2 million currently budgeted for operating the Warehouse itself.

An alternative option, identified above, casts experts in the new FSA Data Office as middlemen in the analysis process. The process would begin with a researcher posing a question or questions to the FSA Data Office, collaborating with system and aid program subject matter experts to refine those questions on the basis of available data. Once the analysis plan was defined, the researcher would author statistical syntax appropriate to the task and provide it to FSA. Staff members tasked to the FSA Data Office would then execute that code against the Warehouse, adjudicating the resulting outputs before returning them to the researchers for their use. Much like our proposal above to expand the utility of the FSA Data Center, we believe the marginal cost associated with this solution is quite low. Although this may require the addition of a small number of new FSA staff, it seems unlikely that annual costs would exceed a few hundred thousand dollars.

Conclusion

As student borrowing rises and concerns about loan debt continue to loom large in the minds of the public, policymakers, and advocates, questions about how students finance postsecondary education will continue to rank high among those charged with providing data, information, and—ultimately— answers. FSA, the owner of the data systems that hold many (if not most) of the keys to better understanding questions about affordability, borrowing, debt, repayment, and default has until recently placed itself in the unenviable position of being the only party in a position to begin to answer them. The examples cited here demonstrate that this need not be the case.

In a data environment where an increasing amount of information is in the hands of qualified analysts, be it through state longitudinal data systems, federated student unit record systems operating on a regional basis, improved use of NCES sample surveys and IPEDS, or further leveraging proprietary data systems like those held by the National Student Clearinghouse, today's limited access to the information housed in FSA's data systems cannot continue. At some point, likely soon, too many stakeholders will argue to too many policymakers that *their* critical question could be answered only if they had better data from the Department. External political pressure, already growing, will become too great to ignore.

Already, there are signs that FSA has realized that governmental researchers outside ED need increased access to their systems, and that there is value too in making these data available to external researchers. The still evolving Enterprise Data Warehouse and Analytics project is tangible evidence of this—but FSA can do more, cheaply, as what are surely thorny questions about who will access the Warehouse (and how) are resolved. And they can be more transparent about acknowledging how increased data access can benefit those trying to improve the student experience and the management of taxpayer dollars.

External pressures for FSA to improve data quality, transparency, and accessibility are unlikely to abate. All signs are that they have already become too great to ignore. FSA, well aware of the situation it finds itself in, must act now to find the most effective, efficient, and privacy-protected ways to make information available to those calling for it, and then let them know their voices have been heard.

Endnotes

1 U.S. Department of Education. (2016). *Student aid administration: Fiscal year 2017 budget request*. Retrieved from http://www2.ed.gov/about/overview/budget/budget17/justifications/y-saa.pdf
2 According to the Fiscal Year 2015 Federal Student Aid Annual Report, "Federal Student Aid has oversight responsibilities for over $1.2 trillion in federal student loans, of which approximately $979.0 billion is directly owned and managed by FSA." Office of Federal Student Aid, U.S. Department of Education. (2015). *Annual report FY 2015*. Retrieved from https://studentaid.ed.gov/sa/sites/default/files/FY_2015_FSA_Annual_Report_official.pdf
3 Ibid.
4 See Looney, A., & Yannelis, C. (2015). *A crisis in student loans? How changes in the characteristics of borrowers and in the institutions they attended contributed to rising loan defaults*. Retrieved from Brookings Papers on Economic Activity website: http://www.brookings.edu/~/media/projects/bpea/fall-2015_embargoed/conferencedraft_looneyyannelis_studentloandefaults.pdf
5 Office of Federal Student Aid, U.S. Department of Education. (2016). *Federal student aid portfolio summary*. Retrieved from https://studentaid.ed.gov/sa/sites/default/files/fsawg/datacenter/library/PortfolioSummary.xls
6 Federal Reserve Bank of New York. (2015). *Quarterly report on household debt and credit*. Retrieved from https://www.newyorkfed.org/medialibrary/interactives/householdcredit/data/pdf/HHDC_2015Q3.pdf
7 Office of Federal Student Aid, U.S. Department of Education (2015).
8 Authors' calculation from U.S. Census Bureau QuickFacts, http://www.census.gov/quickfacts/table/PST045214/00
9 Federal Reserve Bank of New York (2015).
10 Office of Federal Student Aid, U.S. Department of Education (2015).
11 Office of Federal Student Aid, U.S. Department of Education. (2016). *Title IV program volume reports*. Retrieved from https://studentaid.ed.gov/sa/about/data-center/student/title-iv/
12 Ibid.
13 Higher Education Amendments of 1998, Public Law 105–244, 105th Congress (1998).
14 Based on information provided by FSA to the Office and Management and Budget as part of mandatory, yearly IT reporting, FSA has 33 separate IT "business cases," both major and minor. Among the best known are: (1) the Application and Eligibility Determination system, which includes CPS; (2) COD; (3) the Data Challenges and Appeals Solution, through which institutions appeal FSA calculations related to Cohort Default Rates and Gainful Employment; (4) the Debt Management and Collections System; (5) eCampus-Based System, which supports programs such as Perkins Loans and Federal Work-Study; (6) the Enterprise Data Warehouse and Analytics project, described below; (7) NSLDS, described below; (8) the Ombudsman Case Tracking System; and (9) [PEPS, described below.
15 Office of Federal Student Aid, U.S. Department of Education. (2015). *Data extracts*. Retrieved from https://www2.ed.gov/offices/OSFAP/PEPS/dataextracts.html
16 Office of Federal Student Aid, U.S. Department of Education. (2015). *User guides*. Retrieved from http://www2.ed.gov/offices/OSFAP/PEPS/userguides.html
17 Office of Federal Student Aid, U.S. Department of Education. (2015). *The ISIR guide 2016–2017*. Retrieved from https://www.fsadownload.ed.gov/Repository/ISIRGuide1617Nov/2016-2017_ISIR_Guide_FINAL.pdf

18 Office of Federal Student Aid, U.S. Department of Education. (2015). *Technical reference for electronic data exchange (EDE) 2016–2017*. Retrieved from https://www.fsadownload.ed.gov/Repository/EDETechRef1617Nov/2016-2017_EDE_Technical_Ref_Nov2015.pdf

19 Office of Federal Student Aid, U.S. Department of Education. (2015). *Application and verification guide 2015–2016*. Retrieved from https://ifap.ed.gov/fsahandbook/attachments/1516AVG.pdf

20 Office of Federal Student Aid, U.S. Department of Education. (2015). *Common origination and disbursement 2016–2017 technical reference. Volume I: Getting started*. Retrieved from https://ifap.ed.gov/codtechref/attachments/1617CODTechRefVol1.pdf

21 Office of Federal Student Aid, U.S. Department of Education. (2015). *Common origination and disbursement 2016–2017 technical reference. Volume II: Common record technical reference*. Retrieved from https://ifap.ed.gov/codtechref/attachments/1617CODTechRefVol2.pdf

22 Ibid.

23 Dorsey, D. (2005). *Privacy impact assessment for the common origination and disbursement (COD)*. Retrieved from Office of Federal Student Aid, U.S. Department of Education website: https://www2.ed.gov/notices/pia/cod_020305.pdf

24 Office of Federal Student Aid, U.S. Department of Education. (2015). *NSLDS enrollment reporting guide*. Retrieved from http://ifap.ed.gov/nsldsmaterials/attachments/NewNSLDSEnrollmentReportingGuide.pdf

25 Baker, J. (2012). *GEN-12-06: NSLDS enrollment reporting process*. Retrieved from Office of Federal Student Aid, U.S. Department of Education website: https://www.ifap.ed.gov/dpcletters/GEN1206.html

26 Office of Federal Student Aid, U.S. Department of Education. (2015). *NSLDS gainful employment user guide*. Retrieved from http://ifap.ed.gov/nsldsmaterials/attachments/NSLDSGainfulEmploymentUserGuide.pdf

27 Office of Federal Student Aid, U.S. Department of Education. (2015). *National student loan data system school portfolio report extract file layout for schools (SCHPR1) comma separated values*. Retrieved from http://ifap.ed.gov/nsldsmaterials/attachments/NSLDSSchoolPortfolioFileLayoutSCHPR1CSV.pdf

28 U.S. Department of Education. (2016). *Privacy act system of record notice issuances (SORN)*. Retrieved from http://www2.ed.gov/notices/ed-pia.html

29 As described in 5 U.S.C. § 552a(b)(1)-(12), other exceptions include: intra-agency sharing on a need-to-know basis, disclosures required by a Freedom of Information Act request, disclosure to the Census to facilitate a census or other Title 13 activity, required transfer to the National Archives, compliance with a law enforcement request, response to a compelling health or safety need, by order of a Congressional committee, requests from the General Accounting Office, court orders, or activities as described under the Debt Collection Act of 1982.

30 Office of Federal Student Aid, U.S. Department of Education. (2009). Notice of altered system of records. 64 Fed. Reg. 30,173. Retrieved from http://www2.ed.gov/notices/sorn/18-11-09_060499.pdf

31 Office of Federal Student Aid, U.S. Department of Education. (2010). Notice of altered system of records. 75 Fed. Reg. 59,245. Retrieved from http://www.gpo.gov/fdsys/pkg/FR-2010-09-27/pdf/2010-24162.pdf

32 Office of Federal Student Aid, U.S. Department of Education. (2011). Notice of altered system of records. 76 Fed. Reg. 46,778. Retrieved from http://www.gpo.gov/fdsys/pkg/FR-2011-08-03/pdf/2011-19607.pdf

33 Office of Federal Student Aid, U.S. Department of Education. (2011). Notice of altered system of records. 76 Fed. Reg. 37,097. Retrieved from http://www.gpo.gov/fdsys/pkg/FR-2011-06-24/pdf/2011-15747.pdf

34 Ibid.

35 Ibid.

36 The Privacy Act of 1974, Public Law 93–579, 93rd Congress (1974).

37 U.S. Department of Education. (2015). *Family education rights and privacy act (FERPA)*. Retrieved from http://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

38 FERPA would, for example, hold schools responsible for downloading and inappropriately sharing FSA data.

39 U.S. Government Accountability Office. (2005). *Student financial aid: Need determination could be enhanced through improvement in Education's estimate of applicant's state tax payments* (GAO-05-105). Retrieved from http://www.gao.gov/new.items/d05105.pdf

40 Mercer, C. (2007). *Reconfiguring the federal Pell Grant program: Effect of selected changes on program costs and on students in different income groups* (Report RL34084). Washington, DC: Congressional Research Service. Retrieved from https://file.wikileaks.org/file/crs/RL34084.pdf

41 Campbell, C. & Hillman, N. (2015). *A closer look at the trillion: Borrowing repayment, anddDefault at Iowa's community colleges.* Association of Community College Trustees. Retrieved from: http://www.acct.org/files/Publications/2015/ACCT_Borrowing-Repayment-Iowa_CCs_09-28-2015.pdf

42 Looney, A., & Yannelis, C. (2015). *A crisis in student loans? How changes in the characteristics of borrowers and in the institutions they attended contributed to rising loan defaults*. Retrieved from Brookings Papers on Economic Activity website: http://www.brookings.edu/~/media/projects/bpea/fall-2015_embargoed/conferencedraft_looneyyannelis_studentloandefaults.pdf

43 U.S. Department of Education. (2015). *Data documentation for College Scorecard*. Retrieved from https://collegescorecard.ed.gov/assets/FullDataDocumentation.pdf

44 Schneider, M. (2015). *Labor market experiences after postsecondary education: Earnings and other outcomes of Florida's postsecondary graduates and completers*. Retrieved from American Institutes for Research, College Measures website: http://www.collegemeasures.org/file.axd?file=ER_Report.pdf

45 Vaden, E., & Seastrom, M. (2014). *Privacy impact assessment for IES data center (IESDC)*. Retrieved from Office of Institute for Education Sciences, U.S. Department of Education website: http://www2.ed.gov/notices/pia/iesdc.pdf

46 Amazon Web Services. (2016). *AWS GovCloud (US)*. Retrieved from https://aws.amazon.com/govcloud-us/

47 Based on a search of the relevant Department of Education contract, ED-IES-12-C-0094, in the Federal Procurement Data System: https://www.fpds.gov/fpdsng_cms/index.php/en/

48 U.S. Government Accountability Office. (2015). Federal student loans: *Key weaknesses limit education's management of contractors* (GAO-16-196T). Retrieved from http://www.gao.gov/assets/680/673726.pdf

49 U.S. House Committee on Education and the Workforce. (2015). *Foxx statement: Hearing on "Federal Student Aid: Performance-Based Organization Review."* Retrieved from http://edworkforce.house.gov/news/documentsingle.aspx?DocumentID=399777.

50 U.S. Department of Education (2016). Student aid administration: Fiscal year 2017 budget request. Page Y-34. Retrieved from: http://www2.ed.gov/about/overview/budget/budget17/justifications/y-saa.pdf

51 U.S. Department of Education (2016). Student aid administration: Fiscal year 2017 budget request. Page Y-34. Retrieved from: http://www2.ed.gov/about/overview/budget/budget17/justifications/y-saa.pdf

52 U.S. Department of Education. (2014). *Student aid administration: Fiscal year 2015 budget request*. Retrieved from https://www2.ed.gov/about/overview/budget/budget16/justifications/z-saa.pdf

53 U.S. Census Bureau, Department of Commerce. (2015). *U.S. Census Bureau's fiscal year 2016 budget estimates as presented to Congress February 2015*. Retrieved from http://www.osec.doc.gov/bmi/budget/FY16CJ/Census_2016_CJ.pdf

IHEP
INSTITUTE FOR HIGHER EDUCATION POLICY