



Suggestions for Improvements to the Collection and Dissemination of Federal Financial Aid Data



**SUGGESTIONS FOR IMPROVEMENTS
TO THE COLLECTION AND DISSEMINATION OF
FEDERAL FINANCIAL AID DATA**

Report of the National Postsecondary Education Cooperative



This project has been funded, either wholly or in part, with federal funds from the U.S. Department of Education under Coffey Consulting, LLC's Contract No ED-04-CO-0044-0001.

The contents of this publication do not necessarily reflect the views or policies of the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

National Postsecondary Education Cooperative

The National Postsecondary Education Cooperative (NPEC) was established by the National Center for Education Statistics (NCES) in 1995 as a voluntary organization that encompasses all sectors of the postsecondary education community including federal agencies, postsecondary institutions, associations, and other organizations with a major interest in postsecondary education data collection. In 2007, NCES assigned NPEC the responsibility for developing a research and development agenda for the Integrated Postsecondary Education Data System (IPEDS). IPEDS is the core postsecondary education data collection program for NCES. NPEC also occasionally produces products of value to postsecondary data providers, users, and institutional representatives.

NPEC publications do not undergo the formal review required for standard NCES products. The information and opinions published in them are the products of NPEC and do not necessarily represent the policy or views of the U.S. Department of Education or NCES.

October 2011

The NCES Home page is <http://nces.ed.gov>.

The NCES Publications and Products address is <http://nces.ed.gov/pubsearch>.

The NPEC Home Page is <http://nces.ed.gov/npec>.

This publication is only available online. To download, view, and print the report as a PDF file, go to the NCES Publications and Products address shown above.

This report was prepared in part under Contract No. ED-04-CO-0044-0001 with Coffey Consulting, LLC. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Suggested Citation

Steele, P. (2011). *Suggestions for Improvements to the Collection and Dissemination of Federal Financial Aid Data* (NPEC 2012-834). U.S. Department of Education. Washington, DC: National Postsecondary Education Cooperative. Retrieved [date] from <http://nces.ed.gov/pubsearch>.

Content Contact

Elise Miller
(202) 502-7318
Elise.Miller@ed.gov

Table of Contents

Introduction	1
Findings	2
Description and Key Findings by Federal Financial Aid Data Sources.....	3
Recommendations for Improving Financial Aid Data Collection and Dissemination	7
Expand IPEDS Data Collection and FSA and OPE Data Reports to Improve the Granularity and Coverage of Institution-Level Data on Financial Aid Usage	8
NCES Should Provide Leadership of Coordination, Standardization, and Improved Timeliness of All Department Data Reports and Releases	9
Expand the Accessibility of NSLDS Data for Research, and Make Necessary Modifications That Would Allow for the Ongoing Use of NSLDS in the Calculation of Cumulative Debt	10
Implementation of Recommendations	11
References	13
Appendix A: Description of Data Sources Commonly Used to Study Student Financial Aid	15
Appendix B: Detailed Findings Regarding Existing Financial Aid Data Sources	19
Appendix C: Memo Regarding Challenges Matching Campus-Level Data	23
Appendix D: NPEC Financial Aid Data Working Group	33

This page left intentionally blank.

INTRODUCTION

Student financial aid makes college possible for millions of people in the United States. During the 2009-10 academic year, federal and state governments, institutions, and private entities provided almost \$200 billion in financial aid to undergraduate and graduate students in the form of grants, work-study, federal loans, tax credits, and deductions (Baum, Payea, and Cardenas-Elliott (2010). Timely, accurate, and comprehensive information concerning financial aid can help students make more informed decisions about where to go to college and how to finance their education, as well as help educators, researchers, and policymakers understand and address the persistent disparities in access to college opportunity and success in college achievement.

Several offices within the U.S. Department of Education (ED) collect and disseminate data about financial aid. However, limitations of these data sources sometimes make it difficult for consumers, policymakers, and researchers to gain a complete picture of the sources, types, and amounts of aid going to students at institutions of higher education and the relationship between aid and policy goals such as access and success.

The National Postsecondary Education Cooperative (NPEC) is a voluntary organization that includes representatives from federal agencies, postsecondary institutions, associations, and other organizations that have a major interest in postsecondary education data. NPEC's mission is to promote the quality, comparability, and utility of postsecondary data and information that support policy development at the federal, state, and institution levels. In accordance with that mission, NPEC created the Working Group on Financial Aid Data to identify potential improvements to federal data collection and dissemination to address these limitations. This paper presents the Working Group's findings and recommendations for improvements.

The Working Group identified questions that consumers, policymakers, and researchers ask about financial aid. Many of these questions relate to which students receive aid and how much aid they receive, both at a national level, and within sectors, states, or at individual institutions. Other questions address the relationship between financial aid, college access, and completion. Specific questions include:

- How much did the most recent program completers borrow for their education and what type of loans did they take out?
- How many students receive grants and scholarships from colleges and how much do they receive?
- How do students from different backgrounds use financial aid?
- Does receiving aid of different types increase the likelihood of graduation?
- How do patterns of financial aid usage vary from one college to another?
- How have patterns of financial aid usage changed in recent years?
- How do changes in federal financial aid policy affect student outcomes?

The Working Group then looked at the existing federal sources of financial aid data and considered the strengths and limitations of each source in terms of the accessibility and usefulness of the information for answering these questions. The Group reviewed the Integrated Postsecondary Education Data System (IPEDS) Student Financial Aid (SFA) survey instrument, examined how IPEDS data may be complemented with other important financial aid data sources

and considered ways to improve the usefulness and expand availability of existing data sources. The Working Group recognizes that some of the recommendations would require greater reporting burden on institutions; however, others recommendations may decrease it. The group also recognizes that in order to implement these recommendations, the Department will need to determine an appropriate balance between improving access to important financial aid data and increased costs to institutions in providing that information. The Working Group also notes that some of the recommendations will require additional staff and financial resources for ED and its data systems, and that therefore they will vary in the degree of feasibility, particularly in the short term.

In the report that follows, the Working Group identified:

- (1) changes to enhance the usability of IPEDS SFA survey data;
- (2) suggestions for improvements to financial aid data reporting from Federal Student Aid (FSA) and the Office of Postsecondary Education (OPE), with attention to standardization of identifiers, consistency across data releases, and improved ED coordination of data reports; and
- (3) opportunities for better use of existing National Student Loan Data System (NSLDS) data.

FINDINGS

The Working Group reviewed several key sources of financial aid data (see appendix A for a detailed list) and evaluated each source's strengths and limitations for addressing various analytical questions concerning financial aid. The findings presented primarily focus on data in IPEDS, NSLDS, and data files available through FSA and OPE, with brief attention to the National Postsecondary Student Aid Study (NPSAS).¹ To summarize the findings of the Working Group:

- **Some data sources cannot be easily linked or merged by researchers due to inconsistencies in institution identifiers across ED data systems; NCES could best serve to improve coordination of data across offices and enhance data accessibility.** It is difficult to merge data from IPEDS with data from FSA and OPE data reports due to inconsistencies in institution identification codes (see appendix C for a detailed analysis of this issue). Given its experience using OPE and other data on College Navigator and NSLDS in its sample surveys, NCES could play an important role in coordinating the public dissemination of postsecondary education data from Department sources to ensure consistency and standardization of data releases.
- **Some data sources are either too limited or too broad in the student population described and modest changes could provide much more comprehensive information.** Much of the financial aid data in IPEDS focus primarily on first-time/full-time undergraduates, which does not account for the increasingly diverse student

¹ The primary focus of this NPEC Working Group is IPEDS and related financial aid sources. However, several additional federal and nonfederal sources of data are used for analysis of financial aid programs and their effect. See appendix A for a complete list, and appendix B for detailed findings on the strengths and limitations of each source reviewed by the Working Group.

population, including part-time, transfer, and returning students. Data files from FSA and OPE often lump undergraduates and graduate students together, but separate each federal aid program. Collecting additional financial aid data on all students in IPEDS, and disaggregating the level of student and combining data for certain related programs in FSA and OPE data releases, would increase the usability of these data about federal grants and loans.

- **Time lags in the collection and release of data and infrequent periodicity hinder the usability of the data for decision-making.** The National Postsecondary Student Aid Study (NPSAS) is a rich source of detailed financial aid data, but is an expensive data collection that is only collected every 4 years. This limits its usefulness for policy analysis. Due to the nature of financial aid data reporting and that it takes institutions time to account for and report its full distribution to students, IPEDS SFA data are up to 2 years old when they are publicly released. ED could consider the feasibility of increasing the frequency of data collections and/or committing additional resources to decreasing the time between the award year for the data being collected and when they are publicly released.
- **NSLDS is unavailable for addressing policy and research questions.** NSLDS, which collects a rich array of information on all students who receive federal grants or loans, mainly functions as an administrative data source for recording transactions, with limited access and limited use for policy analysis and consumer information.

Description and Key Findings by Federal Financial Aid Data Sources

SFA is one of nine survey components in IPEDS collected by NCES.² IPEDS is administered annually to more than 7,000 postsecondary education institutions in the United States, and institutions that participate in the Higher Education Act Title IV student aid programs are required to report to it under their program participation agreement. IPEDS data provide information for consumers, educators, policymakers, and researchers, and are freely available in multiple formats online. All the data that are collected can be downloaded, and the IPEDS Data Center allows for comparing institutions on a single IPEDS variable, viewing descriptive statistics at the institutional level, or generating reports.

Most questions in IPEDS SFA survey collect data on the number of first-time/full-time degree- and certificate-seeking undergraduate students who receive different types of student financial aid, including grants and loans, from federal and nonfederal sources, as well as the average dollar amount of aid received.³ This first-time/full-time cohort was established to align with the statutorily defined Student-Right-to-Know cohort that is used in the disclosure and reporting of graduation rates by institutions, and to allow for appropriate “apples-to-apples” comparisons across institutions. For example, limits on federal borrowing vary depending upon a student’s year in school which would impact student loan data being reported. Or if institutions reported

² The other IPEDS survey components are Institutional Characteristics (including institutional prices); Fall Enrollment: 12-Month Enrollment; Completions (degrees and certificates conferred); Graduation Rates; 200% Graduation Rates; Human Resources; and Finance.

³ The Higher Education Act, Section 132, added by the Higher Education Opportunity Act of 2008 (20 USC 1015a) specifies data the Secretary must make available to the public on the consumer site, College Navigator (Higher Education Authority Sec. 132(i)(1)(O) – (S)).

on all students and had a large percentage of part-time students, their financial aid packages may seem less generous than an institution basing its awards on mostly full-time students. The SFA cohort ensures first-year students and full-time students are being compared across the institutions. Recent additions to SFA survey include the collection of data to calculate net price figures for first-time/full-time undergraduates. In addition, data are collected for all undergraduates on as well as data on Pell Grants, total grants from all sources, and federal loans.⁴

One important limitation in many IPEDS financial aid questions is that data for the first-time/full-time undergraduate cohort do not necessarily reflect the experiences of returning students, transfer students, part-time undergraduates, and all graduate students. The percentage of first-time/full-time undergraduates receiving a particular type of aid, or the average amount received can be quite different from the equivalent figures for all undergraduates. Some colleges practice front-loading of grants, where college freshmen receive a greater proportion of grants in the financial aid package than upperclassmen. For example, table 1 uses data from NPSAS to show how varying the student population leads to significant differences in the average aid amount at a national level. If analysts intend to examine average aid awards of students, they would obtain very different results using a first-time/full-time cohort of students compared to the total undergraduate population. In most cases the analysis would result in higher average grants and loans for the first-time/full-time students compared to all undergraduate students. On average, first-time/full-time undergraduates with aid received \$2,375 more in grants than the cohort of all undergraduates with aid. Across the entire cohort, first-time/full-time undergraduates received \$2,470 more grant aid than all undergraduates and \$503 more in loans.

Table 1. Average Student Grant and Loan Aid to Undergraduates: 2007–08

	First-time/Full-time undergraduates (\$)	All undergraduates (\$)	Dollar difference (\$)	Percentage difference (%)
Among undergraduates students receiving aid				
Average total grants	7,240	4,865	2,375	49
Average total loans	6,705	7,100	-395	-6
Among all undergraduate students				
Average total grants	4,986	2,516	2,470	98
Average total loans	3,235	2,732	503	18

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2007–08 National Postsecondary Student Aid Study (NPSAS:08).

SFA includes a few questions that cover all undergraduates, but more detailed data are collected only for first-time/full-time students: unduplicated counts of aid recipients and total dollars in each of the following categories: Pell Grants, all federal grants, state/local grants, institutional grants, federal loans, and other (nonfederal) loans. If these same questions were expanded to

⁴ Higher Education Act, Section 132, added by the Higher Education Opportunity Act of 2008 (20 USC 1015a), requires the collection and dissemination of net price data (overall and by income) for the specified subsets of the first-time/full-time cohort; and also in 132(i)(1)(O) and (R), specifies collection of the percentage of first-time/full-time undergraduates receiving different types of aid and the average amounts for the different types of aid.

include information for all undergraduates, full-time undergraduates and part-time undergraduates, users of the data would have a much more comprehensive picture of financial aid for undergraduates. For example, right now, an analyst examining how institutional grant aid varies by institution only has data for first-time/full-time undergraduates. As shown above, this may not accurately represent the experience of the full undergraduate population.

One further limitation of IPEDS SFA is that it shows only annual snapshots of financial aid usage, but no information on cumulative usage. Expanding SFA to include data on cumulative borrowing and cumulative grant receipt would increase the usefulness of these data for understanding students' usage of financial aid as they progress through and complete programs of study.

Currently, SFA data are collected by IPEDS in the spring of the subsequent academic year in which the aid was awarded, and then disseminated through the consumer-oriented College Navigator website and the IPEDS Data Center. Data are made available on College Navigator within 3 months of data collection, and in the IPEDS Data Center for analysts' use within 10 months of data collection. These sites provide comprehensive, but delayed, access to the data, but the full financial aid data are not publicly available until almost 2 years after the end of the award year. The time lag diminishes the utility of this information for consumers and for policy decisions.

Finally, it requires additional time and work for analysts to use IPEDS data in conjunction with other data on financial aid applications and programs from FSA, due to different institutional identifiers. In addition, data do not always align because IPEDS data are reported by institutions at a specific point in the aid cycle, for a specific group of students, whereas FSA data come from the administrative operations such as NSLDS, and may cover a different time period or group of students at the institutions. The resulting matching problems between different ED data files are discussed in appendix C. Not being able to easily use information from these different sources in conjunction with each other hinders analysts' ability to gain insight into the relationship between completion and financial aid usage or outcomes, such as loan repayment patterns.

The ***National Student Loan Data System (NSLDS)*** is a database of information about federal financial aid that is maintained by FSA. NSLDS includes a rich array of data about student financial aid, covering both current and past students, and those who have completed their degrees and those with unfinished degrees. The database tracks federal student loan borrowers and grant recipients from their application for aid, to the receipt of aid and through the loan repayment phase, and also contains information from the Free Application for Federal Student Aid (FAFSA) about their economic and demographic backgrounds and information about the institutions where they enrolled. Students, institutions and lenders have access to certain data from NSLDS to facilitate the administration of aid programs. FSA generates institution-level loan volume and Cohort Default Rate (CDR) data from this source and makes them publicly available on its website. While NSLDS could be a potentially useful source of information and analysis on cumulative debt, loan default, and degree outcomes for students of different economic circumstances, it has been used primarily for administrative purposes and has not generally been available to researchers or analysts. One of the current limitations of NSLDS that hinders these potential uses is the inconsistent reporting of certificate and degree completions. One of the stipulated purposes of NSLDS is to "support research studies and policy development," and a better utilization of the NSLDS data would help ensure that financial aid

programs are developed in a way that best supports college access and completion (Eliades 2007).

NSLDS has longitudinal data about students' usage of several federal aid programs, which could allow for the calculation of cumulative aid usage of students and the analysis of how different aid programs interact. The data on student and institutional characteristics would allow for analysis of financial aid usage for specific populations at different types of institutions. The life-cycle data on student loans could allow for analysis and expanded understanding of student debt accumulation, repayment and default patterns, and how they vary by student and institution types. For example, analysts could use the data to investigate how student characteristics are related to default on federal loans, in order to identify predictors of loan default. Researchers and policymakers could use this information to better target programs that seek to alleviate repayment difficulties by proactively identifying at-risk borrowers before they go into default. Additionally, NSLDS could be used to provide more timely data on students' usage of loans than what is currently available in IPEDS and NPSAS.

The Working Group found several issues that contribute to the underutilization of this source. The publicly available documentation of data elements and data structure was fragmented, outdated, and confusing. It was therefore difficult to understand the content of the data and to discuss potentially beneficial uses for them. The Working Group also found that while at least some data in NSLDS are collected for each campus location (using 8-digit OPE identification numbers to identify campuses), extracted reports are always published for whole institutions (6-digit OPE identification numbers which represent both the main reporting campus and additional campus locations). Institutions have the flexibility to report at either level for some aid programs, so therefore FSA only reports at the 6-digit level. This makes it difficult to merge these data with IPEDS or other sources where data are reported at the campus level, hindering the ability to look at data on financial aid programs alongside data on enrollment or completion. For more information about matching problems between data sources, see appendix C.

Other Federal Student Aid and Office of Postsecondary Education Data Reports provide valuable information with detailed and often current data about various aid programs. These reports include institution-level grant volume data extracts from the Common Origination and Disbursement system; institution-level loan volume data reports with loan volume data by program; institution-level cohort default rates; and FAFSA data reports with applicant volume by state of legal residence and by institution.

One significant weakness of these reports is that data are typically reported by financial aid program, making it difficult to consolidate and analyze data across programs. In almost all cases, data for undergraduate and graduate students are combined, limiting the interpretation of the data. For example, the debt burden to earn an advanced degree is often higher than the debt burden for a bachelor's degree, but the current data reporting does not distinguish between these different student populations. Additionally, the reports on loan volume fail to provide an unduplicated count of the total recipients of unsubsidized and subsidized Stafford Loans, but instead indicate counts for each type of loan. This means that an analyst using these data knows little about the average borrowing per student, but just the average loan by program.

Another shortcoming of this data source is that data files are released to the public or revised without formal public announcements and without sufficient supporting documentation, thus

making it more difficult to use the data to provide meaningful, easily accessed consumer information, or to analyze trends and patterns for policy and research purposes. Additionally, some data are released to the public while the underlying information is still being collected for a given award year, leading to potentially misleading information and significant revisions to the data over time. Existing caveats and warnings are not prominent or clear enough for non-technical users. The data files are not accessible through a centralized location, making it more difficult for analysts, researchers and policymakers to find the data. The FSA Data Center (www.fsadatacenter.ed.gov) has helped improve the accessibility of data, but there are still numerous scattered data resources.

As discussed in more detail in appendix C, improved coordination with NCES and standardization of data reports, identifiers, and releases could greatly expand the public accessibility of these data for analytical use in research and policymaking decisions. Without a common, standard, reliable mapping of institutions and their relationships to campuses and administrative units, and without improvements to data releases, great cost and effort are required to use FSA and OPE data sources together, along with related data from IPEDS, to create the composite of information most meaningful for understanding institution-level financial aid data in context.

The *National Postsecondary Student Aid Study (NPSAS)* is a nationally representative sample survey administered by NCES that covers student enrollment, demographic, and financial aid data. It is the most comprehensive and detailed federal source of financial aid data, but it is only administered once every 4 years. The quadrennial cycle of data collection makes it difficult to use these data for immediate policy discussions or to answer pressing questions that require real-time data, such as an assessment of the effect of the recent credit crisis and economic downturn on financial aid and how students and families pay for college. Because it is a nationally representative sample survey, NPSAS does not yield statistically reliable institution-level estimates or (in most cycles) state-level estimates. This limits its application to consumer needs and certain kinds of research and policy questions, especially those involving the variation in financial aid policies and outcomes across states or institutions. The Working Group did not focus on NPSAS, which is the topic of another ongoing advisory panel. However, some of the Working Group's suggestions would mitigate the infrequency and lack of institution-level data in NPSAS by enhancing the scope and accessibility of annual, institution-level data collections.

RECOMMENDATIONS FOR IMPROVING FINANCIAL AID DATA COLLECTION AND DISSEMINATION

The NPEC Working Group recommends several improvements in the collection and dissemination of federal financial aid data to enhance its usefulness for consumer information, policymaking and research.

First, IPEDS and select data reports from FSA and OPE could be expanded to enhance their usability for addressing financial aid research and policy questions, as well as consumer information on college financing. Second, NCES could provide leadership in the coordination of data reporting and public announcements pertaining to all of the financial aid data released by

ED. Through this coordinating role, NCES could ensure that data are reported with appropriate standardized identifiers to ensure that various data sources can be combined, and that reports meet basic documentation requirements. NCES could also seek out ways to make data available in a timelier manner. Third, ED could explore the feasibility of expanding opportunities for better utilization of existing NSLDS data for research and for alleviating institutional reporting burden.

Expand IPEDS Data Collection and FSA and OPE Data Reports to Improve the Granularity and Coverage of Institution-Level Data on Financial Aid Usage

IPEDS is currently the best source of annual information on financial aid for undergraduates. Through incremental extension of the SFA survey, IPEDS could be used to capture a more complete picture of financial aid usage that reflects the increasingly diverse student population. FSA and OPE could also make incremental changes to enhance existing aid program and FAFSA reports by adding greater detail and disaggregation, and by adding some new detailed reports on loan volume, grants, and FAFSA applications.

1. Expand the scope of IPEDS SFA survey

- Expand the scope of the detailed questions on federal, state, and institutional grants and federal and nonfederal loans to collect data on all undergraduates, all full-time undergraduates and all part-time undergraduates, in addition to the first-time/full-time undergraduate cohort currently covered.
- Add questions to capture data on cumulative debt of students at each level of study (undergraduate certificates, associate's degrees, bachelor's degrees), including both non-completers and graduates. Alternatively, data on cumulative debt could be generated through enhancements to reporting enrollment and completions in NSLDS and combined with existing IPEDS data to alleviate institutional reporting burden.
- Assess the feasibility of disaggregating need- and merit-based institutional grants by facilitating discussions with the financial aid, research and policy community about the most meaningful way to collect and present such data.

2. Provide more detailed loan and grant data reports from FSA and OPE

- At the institution level, enhance existing reports by adding unduplicated counts of Stafford loan recipients, Stafford and PLUS loan recipients together and Title IV grant aid recipients and by disaggregating undergraduate and graduate student borrowing. Current reporting only allows for analysis of average borrowing per program rather than average borrowing per student, because many students take out loans from multiple programs in a given year.
- At the national level, provide reports of historical and cumulative default rates in total dollar amounts and total number of borrowers for more than 3 years, by type of institution. Preserve the historical data for long-term trend analysis.
- Provide new institution-level reports on student loan deferments, forbearances and delinquencies to allow analysis by repayment status.

- Provide a new institution-level report for outstanding loans (borrowers, loans, dollars) by loan program and type. Include institution-level data on choice of repayment plan, especially how it changes with the number of years since entering repayment.
- Offer Pell Grant reports with data on the unduplicated counts of both Pell-eligible students who receive the grant and those who apply but are not awarded the grants.
- Improve reporting of degree completions in FSA data, then calculate and report cumulative federal debt for degree completers by degree type for each institution.
- Disaggregate data for undergraduate students and graduate students in all reports.

3. Expand detailed reporting on FAFSA applications

- Compile information by high school on the number and percentage of students filling out FAFSA forms, and collect more information on students who applied for aid but did not enroll in college.
- Add unduplicated national counts of students filing FAFSA each quarter and each year to existing FAFSA reports.
- Add a count of FAFSA filers who enroll and the number who receive federal aid, to the existing report on the number of FAFSAs filed at each institution.

NCES Should Provide Leadership of Coordination, Standardization, and Improved Timeliness of All Department Data Reports and Releases

Inconsistent reporting methods and insufficient documentation in FSA and OPE financial aid reports contribute to the inaccessibility and incompatibility of some financial aid data, making it difficult to use in research and policy analysis, as well as in consumer information. These reports could be much more accessible with attention to standardization of identifiers, consistency in data releases and improved ED coordination of all data reports related to financial aid. See appendix C for more detail about data matching problems between different ED sources.

4. Improve coordination of data report releases

- All FSA and OPE reports should include enrollment figures that correspond to the same level of reporting as the financial aid data (i.e., campus or system of campuses).
- Clearly document the date on which data were extracted or collected and whether the data might be updated in the future.
- Do not release data that are so preliminary that it vastly undercounts financial aid recipients or dollars for a reporting period.
- Adopt common rules and practices for adding, deleting, and changing the identifier for a given campus or school as changes in status, control, or affiliation occur over time.
- Include a consistent sector designation that is based on program or degree level and type of control in all data systems. Currently, some data reports only differentiate schools by control (public, private nonprofit, private for-profit), while others additionally differentiate schools by institution level (4-year, 2-year, less-than-2-year).
- In cases where it is feasible, report financial aid data based on the attended campus/location (rather than the main reporting campus for a group of schools) for all FSA, OPE, and NCES data.

5. Standardize institutional identifiers to address data matching issues

- Adopt common identifiers for each school and campus/location to be used across all datasets so that data from different sources can be used together.
- Generate all data reports using the 8-digit OPEID.
- For each data year (the academic year that a dataset describes), provide a crosswalk mapping each 6-digit and 8-digit OPEID in FSA and OPE data to the corresponding UNITID(s) and vice versa.
- Include documentation with each dataset detailing which crosswalk to use for that dataset and any caveats or special notes.
- Clearly document any cases where data reported in one file represent more than one campus.
- Provide the 6-digit OPEID in FSA and OPE files where data are reported only at the 6-digit OPEID level, not the 8-digit OPEID for the main reporting campus.
- Format OPEIDs as text so leading zeroes appear, or make IDs consistent in all sources.

6. Consider improvements to timeliness of data releases

- Assess the feasibility of collecting and releasing IPEDS SFA data sooner for research and policy analysts.
- Assess the feasibility of a limited NPSAS sample in between the 4-year time period, to gather more timely data on pertinent policy questions. For instance, this might be addressed by analyzing more frequent data extracts from NSLDS beyond the NPSAS data collection cycle currently used.

Expand the Accessibility of NSLDS Data for Research, and Make Necessary Modifications That Would Allow for the Ongoing Use of NSLDS in the Calculation of Cumulative Debt

Despite containing a wealth of information about financial aid, NSLDS is currently underutilized due to limited access. Data from NSLDS could help policymakers and researchers better understand cumulative debt accumulation for those who do and do not complete degrees, how students manage debt, and how they utilize the various types of deferment and repayment options. Detailed analysis could also identify specific debt issues for low-income students who also receive Pell Grants.

7. Seek expanded analytical uses for NSLDS

- Improve collaboration between FSA, NCES, and others to explore and implement specific consumer, policy, and research uses of NSLDS data, such as providing cumulative grant and loan amounts and percentages with debt on consumer-facing web pages, analyzing student debt loads for those completing degrees and those who do not complete degrees, and analyzing federal grant and loan usage and repayment patterns for specific student populations such as low-income students.
- Make complete and up-to-date documentation of data elements and data structure available publicly.
- Add data and reporting capabilities for nonfederal student loans to NSLDS for use in research and analysis. Currently, very limited annual data are available on nonfederal loans, making it challenging to analyze overall borrowing patterns.

8. Take steps toward data releases of NSLDS

- Extract a sampled and anonymized snapshot of NSLDS data and provide it to NCEES for documentation of data quality and availability issues for consumer, policy, and research applications. This snapshot should include at least a random sample of records with all variables from the NSLDS database. Periodic snapshots could be provided for trend analysis.
- Preserve historical enrollment and completion status information and make it accessible to data users.
- Provide institutions with reports that include both current status and historical data at the aggregate and individual level for their students/borrowers over time. Currently, NSLDS reporting for institutions only provides a snapshot of current status at the institution level and historical data at the individual level.
- Capture degree level (e.g., certificate, associate's, bachelor's) when enrollment and completion status is updated and ensure accurate reporting of completion dates and enrollment status change dates.
- Create a web-based interface to provide access to NSLDS data for research and policy analysis, using a tool similar to NCEES PowerStats with comparable public accessibility.

IMPLEMENTATION OF RECOMMENDATIONS

NPEEC created the Working Group on Financial Aid Data to identify areas for improvement to federal data collection and dissemination. The Working Group approached this task by first analyzing the key questions that consumers, policymakers, and researchers ask about financial aid. The final recommendations were generated by asking two overarching questions:

- (1) What do we need to know about financial aid that we cannot know from existing data?
- (2) How might existing data sources be enhanced to better address current questions about financial aid?

The Working Group weighed the feasibility of the recommendations set forth, and where information was available; we discussed which recommendations seemed achievable in the short-term and long-term.

The Working Group recognizes that some of the recommendations, while important to the improvement of data usability, present distinct challenges and require significant resources. The Working Group recommends that ED evaluate the feasibility of each recommendation in order to assess time and cost for implementation, as well as the potential burden on institutions and other stakeholders involved in data collection.

This page left intentionally blank.

REFERENCES

- Baum, S., Payea, K., and Cardenas-Elliott, D. (2010) *Trends in Student Financial Aid, 2010*. New York: College Board. Retrieved from http://trends.collegeboard.org/downloads/Student_Aid_2010.pdf.
- Eliades, P. (2007). *Privacy Impact Assessment for the National Student Loan Data System (NSDLS)*. U.S. Department of Education. Washington, DC: Federal Student Aid. Retrieved from <http://www2.ed.gov/notices/pia/nslds.pdf>.

This page left intentionally blank.

APPENDIX A:
DESCRIPTION OF DATA SOURCES COMMONLY USED TO STUDY
STUDENT FINANCIAL AID

Federal Data Sources

Integrated Postsecondary Education Data System (IPEDS) Student Financial Aid Component (SFA). IPEDS is a set of surveys conducted annually by the U.S. Department of Education's National Center for Education Statistics (NCES). Information is collected in seven areas: institutional characteristics, institutional prices, enrollment, student financial aid, degrees and certificates conferred, student persistence and success, and institutional human and fiscal resources. IPEDS data provide information for consumers, policymakers, and researchers and is freely available in multiple formats online. Specific survey results can be downloaded as a zipped file (SAS, SPSS, etc.); an online IPEDS data tool allows manipulation of IPEDS data, such as comparing institutions on a single IPEDS variable or generating institution reports; and College Navigator displays descriptive statistics at the institutional level. The new State Data Center provides IPEDS data aggregated by state and other state and national aggregations of IPEDS data can be found in other NCES publications, such as the *Digest of Education Statistics*.

The Student Financial Aid (SFA) survey component is collected annually from institutions for undergraduate students, with data regarding federal grants, state and local government grants, institutional grants, and loans. Data collected include the number of students receiving each type of financial assistance, and average amount received by type of aid. Data are also collected for the calculation of average net price, in accordance with the Higher Education Opportunity Act of 2008. These data are available on the College Navigator website and in the IPEDS data tools mentioned above.

National Student Loan Data System (NSLDS). Maintained by the U.S. Department of Education's Federal Student Aid (FSA), the NSLDS is a comprehensive national database of information about the federal financial aid history of recipients of student financial assistance authorized under Title IV of the Higher Education Act of 1965, as amended. As the central database for Title IV student financial aid, the NSLDS stores information about loans, grants, students, borrowers, lenders, guaranty agencies, schools, and loan servicers. It provides an integrated view of Title IV loans and grants from aid approval through disbursement, repayment, default, and closure. The three main goals of NSLDS according to the user documentation are to (1) improve the quality and accessibility of student financial aid data; (2) reduce the administrative burden of Title IV aid; and (3) minimize abuse within aid programs by accurately tracking funds awarded to assist postsecondary students.

Currently the U.S. Department of Education does not make NSLDS data broadly available for analysis. Prior research and policy studies have used extracts, and the Department uses the data in program reporting, budget forecasts, and Cohort Default Rate (CDR) analyses. However, more detailed information on individual student borrowing aggregated by institution is needed to inform policy and to understand more clearly the financial health of a particular institution's borrowers. Most importantly, there is currently no reliable way to review cumulative borrowing for graduates on an annual basis; however, this could be provided, at least for federal loans, using NSLDS data. Furthermore, for the sources that do gather limited annual cumulative debt

data, none include data on those who never complete their degree. To obtain cumulative borrowing data for both graduations and non-completers, NSLDS would need to require institutions to accurately report both the date and level of completion for each degree or certificate completer.

Among the many purposes identified in the establishment of NSLDS, one stipulated purpose is to support research studies and policy development. The Higher Education Authority mandates research and policy analysis in addition to administrative uses of NSLDS. Data from NSLDS databases have not been structured for research purposes or made widely available for research. Research and analysis is best supported by non-volatile (e.g., unchanging) extracts that represent the status of the system for specific points in time or time periods.

National Postsecondary Student Aid Study (NPSAS). NPSAS is a national sample survey of student enrollment, demographic and financial aid information, sponsored by NCES. It contains student-level information on grants, loans, work-study and other forms of financial aid. Information is collected via institutional records, government databases, and student interviews. Data are available for academic years 1986-87, 1989-90, 1992-93, 1995-96, 1999-2000, 2003-04, and 2007-08 (with the next NPSAS planned for 2011-12). Each data collection contains a nationally representative sample of students from a stratified sample of postsecondary education institutions.

NPSAS is also the source of the initial cohorts for the Beginning Postsecondary Students (BPS) and the Baccalaureate & Beyond (B&B) studies:

BPS is a longitudinal survey tracks a cohort of students drawn from NPSAS sample. Students are surveyed 2 years and 5 years after their initial enrollment in postsecondary education; the information gathered includes student persistence in and completion of postsecondary education programs, their transition to employment, demographic characteristics, and changes over time in their goals, marital status, income, and debt.

B&B is a longitudinal survey study tracks a cohort of students among graduating seniors in the NPSAS sample in each academic major area. The information gathered includes bachelor's degree recipients' undergraduate experience, demographic background, expectations regarding graduate study and work, participation in community service, workforce participation, income and debt repayment and entry into and persistence through graduate school programs. It also has a specific focus on elementary and secondary teachers, specifically tracking teacher preparation, entry into and persistence in the profession and teacher career paths.

Publicly available data from NPSAS, BPS, and B&B are organized into specific data file releases and made available through a web-based Data Analysis System, PowerStats and QuickStats tools. In addition, NPSAS data are archived by and available through the International Archive

Data Beyond Federal Sources

Common Data Set (CDS). The CDS is a joint initiative among three publishers, College Board, Peterson's, and *U.S. News & World Report* and the educational community to standardize the definitions used in higher education related-data. It is intended to be a set of standards for data collection rather than a survey instrument or database. Minor revisions occur annually to ensure continuous improvement to the survey questions.

The CDS questions include information regarding institution type, enrollment, and persistence by demographic characteristics; graduation rates by annual cohort; admissions requirements and statistics; freshman profile; admissions policies such as early action, transfer admission statistics, academic offerings; tuition and other annual expenses; financial aid information and policies; instructional faculty and class size; and disciplinary areas of degrees conferred. CDS publishers provide some data on their websites and may license other datasets to outside users.

National Student Clearinghouse (NSC). NSC is a non-profit organization that manages a database of students with the following information: name, birth date, Social Security number, high school attended, high school diploma date, college enrollment institution and date, transfer history, transcripts for institutions attended, degrees awarded with dates, loan applications and awards, and enrollment status for family members listed on loan applications. NSC databases are filled with information provided by participating high schools, colleges and universities, guarantors, lenders and loan servicers. Approximately 92 percent of U.S. colleges and universities participate in the database and it has information on more than 100 million current and former students.

The NSC provides information on their website for the following purposes:

- Students can access information about their own enrollment and loan status and history.
- High schools can track enrollments and degrees for their graduates.
- Lenders, guarantors, and servicers can track enrollment information for their borrowers
- Companies or individuals who need to verify enrollment, degree status, or certification status for an individual student can do so for a fee.
- Colleges and universities can access student data in individual or aggregate form. If the institution participates in the NSC, it can use the StudentTracker data tool to retrieve information on cohorts of students rather than individuals.

The data retrieval services available to colleges and universities would support NSC database research, but affiliation with a participating institution or a contractual agreement is necessary.

National Association of College and University Business Officers (NACUBO). NACUBO is organization focused on the financial management of colleges and universities. It collects, analyzes, and publishes data in three main areas: (1) endowments, (2) tuition discounting, and (3) student financial services benchmarking.

- NACUBO-Commonfund Endowment Study collects institutional survey data on the management of college and university endowments and affiliated foundations, including endowment size, asset allocation, investment performance, management expenses, spending rates and policies, and institutional debt. Data have been collected annually

since at least 2002. Online access is available for purchase and preview of selected tables is available for free online.

- The annual tuition discounting survey of selected institutions has been conducted for about 18 years. Results are available for free to member institutions or for purchase online. The results files provide a PDF of aggregate institutional data, in the form of tables summarizing tuition discount rates by institutional characteristics.
- NACUBO has an online benchmarking tool that is available only to the institutions that participate in the tuition discounting survey.

APPENDIX B:

DETAILED FINDINGS REGARDING EXISTING FINANCIAL AID DATA SOURCES

Institutional Postsecondary Data System (IPEDS) Student Financial Aid (SFA) Survey		
Overview/Strengths	Limitations	Implications
<ul style="list-style-type: none"> • One of seven areas covered by IPEDS: institutional characteristics, institutional prices, enrollment, student financial aid, degrees and certificates conferred, student persistence and success, and institutional human and fiscal resources. • Data are collected annually, providing important information about yearly trends for undergraduate students. • Institution-level data are gathered that can be aggregated to state, national, and other groupings (though multistate institutions may be counted in their primary state). • Survey includes all institutions that participate in Title IV financial aid programs and other institutions that report voluntarily. • Survey includes a count of all students rather than a sample of students. • Survey gathers grant and loan data from federal, state, and institutional sources. 	<ul style="list-style-type: none"> • The cohort, which is the focus of the survey, of first-time/full-time undergraduates does not necessarily constitute a representative sample of all undergraduates at many institutions. • Limited data are gathered on all undergraduates, and those data do not differentiate between the various sources of the grant dollars. • The annual snapshot of data does not provide sufficient information on the cumulative usage of aid by students (such as total loans borrowed or total grants received throughout a degree program). • There is an almost 2-year time lag between the end of an academic year and public release of data about that year. • It is difficult for many users to generate aggregate figures or comparisons over time. • Combining data from IPEDS with data from FSA is greatly complicated by the use of inconsistent school identifiers. 	<ul style="list-style-type: none"> • Because of the narrow focus, IPEDS SFA provides an incomplete picture of aid use and aid dollars, especially institutional aid. • IPEDS cannot be used to look at cumulative debt, which may be useful for consumers making decisions about college finances. • The time lag diminishes the utility of this information for consumer and policy decisions. • The difficulty of using FSA, OPE, and IPEDS data together hinders researchers' and policymakers' ability to get a comprehensive picture of financial aid and how it interacts with other student- and institution-level factors to influence college access and completion.

National Student Loan Data System (NSLDS)

Overview/Strengths	Limitations	Implications
<ul style="list-style-type: none"> • NSLDS is a comprehensive national database of information about federal financial aid that tracks individual students' grant and loan transactions. • NSLDS includes both current and past students, and those who have completed their degrees and those with unfinished degrees. • Longitudinal data in NSLDS can be used to analyze cumulative loan usage at the student level. • Because data are collected on federal grants, loans, and enrollment for all Title IV aid recipients, NSLDS may be used to analyze interaction between some aid programs. • Student-level data in NSLDS may be aggregated to institution, state and national levels. • The data include detailed information on student and institutional characteristics. • NSLDS is continuously updated, and could be used to provide useful data extracts and reports annually or quarterly. 	<ul style="list-style-type: none"> • Data in NSLDS are not structured for research purposes and there is no public access to data in NSLDS beyond selected extracts. • Publicly available documentation of data elements and data structure is old, fragmented, and difficult to understand. • There is no clear point of contact for discussion on the potential research uses of NSLDS. • In some NSLDS data, 8-digit identification numbers are used, but extracts usually use 6-digit identification numbers, making it difficult to properly merge the information with IPEDS data for analysis. • Publicly available data extracts lumps together undergraduates and graduate students, making it inadequate for meaningful analysis. • Reports from this data source do not allow schools to distinguish between types of deferments, or look at repayment behavior after consolidation. 	<ul style="list-style-type: none"> • NSLDS has the potential to be a unique data source with important research, consumer, and policy applications, and it is currently being underutilized. • With some enhancements to reporting of enrollment and completion status, NSLDS could provide annual data on cumulative debt of completers by degree level, or noncompleters, by institution. • These data could be used to inform policymakers and consumers about student loan borrowing trends, and default and repayment patterns of low-income students compared to those of other students. • Default prevention could be implemented with NSLDS by identifying borrowers at risk and providing counseling. Currently institutions cannot identify variations on deferment or repayment after consolidation. • Greater understanding of debt burden could be obtained from NSLDS with analysis of how students use the different options for deferment and forbearance. • Usage of available repayment plans can be estimated using payment amounts or could be tracked directly with additional variables. • Public abstracts of data about institutions could be used for predicting default by looking at borrowers in a cohort to determine whether changes occur from one fiscal year to the next.

National Postsecondary Student Aid Study (NPSAS)		
Overview/Strengths	Limitations	Implications
<ul style="list-style-type: none"> • Each cohort contains a nationally representative sample of students from a stratified sample of postsecondary education institutions. • It contains student-level information on grants, loans, work-study, and other financial aid, collected via institutional records, government databases, and student interviews. • Sector-level data analysis is available. 	<ul style="list-style-type: none"> • Data are only gathered once every 4 years. • Only a limited number of states were oversampled to obtain a state-level representative sample. • Institution-level data are not available. 	<ul style="list-style-type: none"> • No institution-level analysis can be conducted; thus, the data cannot be used to explore how financial aid usage and outcomes vary by college. • A fairly significant delay occurs between data collection and data availability, making the data less useful for immediate policy questions and concerns (such as changes in borrowing patterns and payment behavior resulting from the recession).

Data Reports From the Federal Student Aid (FSA) and Office of Postsecondary Education (OPE)		
Overview/Strengths	Limitations	Implications
<ul style="list-style-type: none"> • FSA and OPE provide a number of timely annual or quarterly reports for various programs with very detailed information by aid program. • These reports include Grant volume data extracts from the Common Origination and Disbursement system (COD), including the Federal Pell Grant, Academic Competitiveness Grant (ACG), and National Science and Mathematics Access To Retain Talent Grant (National SMART Grant), Federal Pell Grant and Teacher Education Assistance for College and Higher Education Grant (TEACH Grant) programs. • Loan Volume Data reports are provided by program, and the source data are from NSLDS. • Cohort Default Rate data are provided for each federal fiscal year, and these reports include data for all colleges participating in Title IV loan programs and colleges subject to loss of loan program eligibility due to default rates. • FAFSA Data by State and College report on the unique number of applications by state of legal residence and the number of FAFSA forms received at an institution. 	<ul style="list-style-type: none"> • These data sources typically lump undergraduates and graduate students together, making it difficult to do any meaningful analysis. • Generally, the data extracts from these sources are reported with 6-digit OPE identification numbers, making it extremely difficult to merge the information with other sources. • Data are often released without any analytical findings or context and without public announcement. • Data are reported for a single program (or a small group of programs) and lack basic contextual information such as institutions' enrollment figures. • Existing data extracts of these sources are revised after initial posting, but communication about such changes are not explicit or sufficiently user-friendly. • Extracts of loans in these data sources do not provide an unduplicated count of the total recipients of unsubsidized and subsidized Stafford Loans, but instead indicate counts for each type of loan. • There is no clear point of contact for questions regarding these data on the website where information is published. 	<ul style="list-style-type: none"> • Researchers find that the process is extremely time-consuming to compile data for multiple years, for aggregate groupings such as states and sectors, or across different programs; as a result, this information is under-utilized. • Information on the number of students using programs is obscured by lack of an easy means of comparing program usage data to enrollment data. • The difficulty of using FSA, OPE, and IPEDS data together hinders researchers' and policymakers' ability to get a comprehensive picture of financial aid and how it interacts with other student- and institution-level factors to influence college access and completion. • Due to duplicated counts of borrowers, it is not possible to calculate the average borrowing per student, only the average loan by program.

APPENDIX C:

MEMO REGARDING CHALLENGES MATCHING CAMPUS-LEVEL DATA

DATE: June 21, 2010

TO: Elise Miller, IPEDS Program Director; Thomas Weko, Associate Commissioner, Postsecondary Studies Division

FROM: Matthew Reed, Program Director, The Institute for College Access & Success; Chair, NPEC Financial Aid Data Working Group

RE: Challenges in Matching Campus-Level Data from U.S. Department of Education sources

Timely, accurate, and comprehensive data are critical to identifying and addressing what is and is not working in higher education. Consumers, policymakers, and researchers all need user-friendly access to meaningful data – data that can deepen our understanding of important issues and inform decision-making at all levels. However, within the U.S. Department of Education, different datasets use different identifiers for schools and campuses, and some datasets report data for each campus while others group related campuses together. This makes it difficult if not impossible to compare schools or campuses to analyze the impact of federal, state, and institutional policies on outcomes such as enrollment, persistence, and completions. Although other types of data are available, campus-based analyses can uniquely shed light on the relationship between federal, state, and institutional financial aid investments and outcomes for students with varied demographic characteristics.

This memo discusses preliminary findings and recommendations of the NPEC Financial Aid Data Working Group (see Attachment A) regarding matching campus-level data from Department of Education sources. It first describes the major data-matching challenges facing users of these different data sources and their implications for research, policy and practice. It then provides recommendations for improving the usability and accuracy of campus-level data.

The Context

Several agencies within the U.S. Department of Education provide higher education data. The National Center for Education Statistics (NCES) conducts the National Postsecondary Student Aid Survey (NPSAS) every 4 years, which provides representative data at the national level and for several states. More frequently released data at the institution, state, and national levels are available from NCES' Integrated Postsecondary Education Data System (IPEDS) surveys and programmatic data and reports from Federal Student Aid (FSA) and the Office of Postsecondary Education (OPE).

To make full use of these data sources, it is often necessary to combine FSA/OPE data on the usage of specific federal financial aid programs with comprehensive IPEDS data on enrollment, completions, graduation rates, and other topics. For example, determining the percentage of all undergraduates at a school applying for federal aid or using Pell Grants, PLUS loans, or any other financial aid program requires combining FSA program data with IPEDS enrollment data.

The necessity of matching data from different sources in an accurate and comprehensive manner is especially apparent in the Department’s own efforts to combine data from NCES and FSA/OPE sources, both for research and policy purposes *and* for public-facing consumer information. For example:

- Graduation, transfer, and retention rates from NCES’s IPEDS appear on FSA’s FAFSA-On-The-Web (FOTW) site alongside school information from FSA sources.
- Loan volume and cohort default rate data from FSA’s NSLDS appear on NCES’s College Navigator site alongside data from IPEDS.

Without resolving the matching problem, consumer information from the Department may be incomplete or even misleading for certain schools or campuses. For example, a student using FAFSA-On-The-Web to apply for financial aid at the University of Phoenix would not be able to view graduation, transfer, and retention rates for a specific campus due to mismatched levels of reporting; the University of Phoenix only has one entry in FSA and one federal school code, but reports its IPEDS data as 73 separate campuses.

The Root of the Problem: Inconsistent School Identifiers

Within the Department of Education, different data sources use different identifiers for schools and campuses. IPEDS uses a unique 6-digit identifier (UNITID) for each campus included in its surveys.¹ All schools receiving Title IV aid funds are required to report data in IPEDS, and other postsecondary institutions have the option to report.

FSA and OPE data use a separate identifier called the OPEID in their databases, though some databases use other identifiers which are mapped to OPEIDs. At least in the datasets available on the FSA and OPE web sites, the main campus of a school will have a 6-digit OPEID with a 2-digit suffix of “00,” while other campuses will have the same 6-digit OPEID as the main campus, but with a different 2-digit suffix (also called a location code²). The IPEDS Header (HD) and Institutional Characteristics (IC) files include 8-digit OPEIDs (which combine the 6-digit school identifier and the 2-digit location code) in addition to the UNITIDs.

In the FSA Data Center, most datasets have 6-digit OPEIDs or 8-digit OPEIDs where the 2-digit branch code is always “00”, meaning data are effectively reported at the 6-digit OPEID level. It is unclear if this is due to limitations in how the data are collected and stored in the underlying databases, decisions about how to extract and report the data on the public data center, or some combination of both. Data are reported at the 8-digit OPEID level, at least for some schools, in the FSA data file on the usage of Pell Grants and other federal financial aid grants.

While some schools process financial aid centrally across multiple campuses, in many cases it should be possible for the FSA databases to capture information about financial aid usage based on the campus the student attended while receiving the aid (“attending school”), as well as the main office or location of the

¹ In this memo, we use “campus” to mean the institutions with unique UNITIDs in the IPEDS HD file, which generally correspond to the school “locations” in FSA/OPE data. There may be cases where distinct UNITIDs represent different parts of a school that share a physical campus or where one UNITID represents parts of a school that are not physically contiguous. As noted in the recommendations section of this memo, further documentation of the existing definitions for reporting units in both FSA and IPEDS data would be helpful.

² Some “locations” appear to be designated as “branches” while others are not.

school (“reporting school”). This would make it easier to compare data across campuses and combine financial aid usage data with other information about enrollment and completion.

Data-Matching Challenges

To illustrate some of the challenges in matching IPEDS data to FSA/OPE data, this section will provide examples using the FSA grant volume data extracts available on the FSA Data Center,³ the FISAP data file from OPE⁴ and the header (HD) and 12-month enrollment (EFFY) files available on the IPEDS Data Center⁵ for 2007-08. Similar issues arise when using other FSA/OPE data including the institution-level data files on loan volume, campus-based aid, FAFSA applications, and cohort default rate data, as well as the OPE Pell data files.⁶ In contrast, other IPEDS data files can be easily matched to the IPEDS HD file via the UNITIDs.

Incomplete Matches Based on 8-digit OPEID

Even when FSA data are reported with 8-digit OPEIDs, one-to-one matches to IPEDS may not be correct. As background, the IPEDS HD file contains 7,055 entries for postsecondary institutions, each with a unique UNITID and in most cases a unique 8-digit OPEID as well.⁷ Most FSA data files contain entries with unique 8-digit OPEIDs. To merge an FSA file with IPEDS data, most users would start by matching the records based on the 8-digit OPEIDs available in each file. However, some of these matches may not be accurate: an entry in the FSA data extract with a given 8-digit OPEID may not refer specifically to the entry in IPEDS with that same 8-digit OPEID.⁸

To illustrate this point, the FSA data extract on federal financial aid grants has 5,505 entries with unique 8-digit OPEIDs. Merging this FSA file with IPEDS using 8-digit OPEIDs results in 5,353 matches. However, these include *457 cases where the number of Pell Grant recipients reported in the FSA Grant data is greater than the 12-month enrollment reported in IPEDS*. This suggests that these entries in the FSA data extract may actually represent data for more than one location at a school.⁹

For example, the 8-digit OPEID 00467300 matches to one entry in the FSA Grant file for “Baker College” with 22,591 Pell Grant recipients. That 8-digit OPEID matches the campus named “Baker College of Flint” in IPEDS, with a 12-month undergraduate enrollment of only 7,480. This discrepancy suggests that this entry in the FSA Grant file actually represents many if not all of the 11 Baker College

³ See “Federal Student Aid - Data Center – Programmatic Volume Reports,” <http://federalstudentaid.ed.gov/datacenter/programmatic.html>.

⁴ The FISAP is the application schools file with the Department annually to participate in campus-based financial aid programs. The FISAP data file, available upon request from OPE, contains more extensive data than that shown in the campus-based aid data extract posted on the FSA Data Center.

⁵ See “Integrated Postsecondary Education Data System Data Center,” <http://nces.ed.gov/ipeds/datacenter/>.

⁶ See “Federal Student Aid - Data Center – Programmatic Volume Reports,” <http://federalstudentaid.ed.gov/datacenter/programmatic.html>, “Federal Student Aid - Data Center – Application Volume Reports,” <http://federalstudentaid.ed.gov/datacenter/application.html>, “Federal Student Aid – Data Center – Default Rates,” <http://federalstudentaid.ed.gov/datacenter/cohort.html>, “Federal Pell Grant Program End-of-Year Report and Other Annual Data Reports,” <http://www2.ed.gov/finaid/prof/resources/data/pell-data.html>.

⁷ There are 44 cases where the OPEID is missing and 4 cases with non-unique 8-digit OPEIDs.

⁸ This matching problem is evident in FSA files where 8-digit OPEIDs all end in “00” as well as FSA files where some 8-digit OPEIDs have other suffixes.

⁹ There may be additional cases where one entry in the FSA data corresponds to multiple entries in IPEDS. These occur when the 6-digit school identifier (first 6 digits of the OPEID) appears in only one entry in the FSA data but in multiple entries in the IPEDS data, regardless of whether the Pell count is greater than the enrollment.

campuses listed in IPEDS with the 6-digit OPEID 004673, as listed below.¹⁰ However, *there is no definitive documentation of which campuses in IPEDS correspond to each entry in FSA data.* As discussed in the recommendations section, a comprehensive crosswalk between OPEIDs in FSA data and UNITIDs in IPEDS data would be very helpful in ensuring that all entries on both sides are accurately matched, whether the matches are one-to-one, or one-to-many.

FSA Federal Grant Volume File, 2007-08

OPEID	Name	State	Zip Code	School Type	Federal Pell Grant Recipients
00467300	Baker College	MI	48507-5508	Private-nonprofit	22,591

IPEDS HD/EFFY Files, 2007-08

8-digit OPEID	UNITID	Institution Name	State	Zip Code	Sector	12-month undergraduate enrollment
00467300	168847	Baker College of Flint	MI	48507-9987	Private not-for-profit, 4-year or above	7,480
00467301	404648	Baker College of Cadillac	MI	49601-9600	Private not-for-profit, 4-year or above	2,444
00467302	404082	Baker College of Clinton Township	MI	48035	Private not-for-profit, 4-year or above	7,569
00467303	171298	Baker College of Muskegon	MI	49442	Private not-for-profit, 4-year or above	6,708
00467304	168838	Baker College of Owosso	MI	48867	Private not-for-profit, 4-year or above	3,804
00467305	404073	Baker College of Auburn Hills	MI	48326-2642	Private not-for-profit, 4-year or above	5,357
00467306	381617	Baker College of Port Huron	MI	48060	Private not-for-profit, 4-year or above	2,271
00467309	414160	Baker College of Jackson	MI	49202-1290	Private not-for-profit, 4-year or above	2,430

¹⁰ Some campuses may not be represented in the FSA file because they do not participate in campus-based aid, but it is not possible to determine that from the information in these two files alone.

00467313	414708	Baker College Center for Graduate Studies	MI	48507	Private not-for-profit, 4-year or above	7,068
00467317	444167	Baker College of Allen Park	MI	48101	Private not-for-profit, 4-year or above	3,249
004673A2	419572	Baker College Corporate Services	MI	48326-2642	Private not-for-profit, 4-year or above	707

Arbitrary Groupings of Campuses based on 6-digit OPEID

In the example of Baker College (previous page), one entry in an FSA file appears to correspond to a group of entries in IPEDS that share the same 6-digit OPEID. There are also cases where it appears that a school has multiple entries in an FSA dataset, each with its own distinct 6-digit OPEID, corresponding to one or more IPEDS entries.

For example, Vatterott College appears under four distinct OPEIDs in the FSA Grants File, giving the impression that these are four separate schools with the same name or four campuses of the same school. However, looking at the IPEDS file, these same four 6-digit OPEIDs cover 17 campuses in groupings of one to ten campuses per 6-digit OPEID. To illustrate, the percentage of undergraduates with Pell Grants is calculated below, first by matching the full 8-digit OPEID in each file and then by matching each FSA Grant File entry to the IPEDS entries with the same 6-digit OPEID. As in the examples on the previous page, there are *several cases where matching on the 8-digit OPEID is clearly inaccurate, as the number of Pell recipients greatly exceeds the enrollment*. Matching on the 6-digit OPEIDs, we get more plausible figures for the percentage receiving Pell Grants, but they are difficult to interpret. *For the 6-digit OPEID 025997, the Pell Grant percentage of 73% includes many campuses in Missouri but not all the Missouri campuses of the school, and also includes campuses in several other states.*

FSA Grants File, 2007-08

8-digit OPEID	6-digit OPEID	School	State	School type	Pell Grant recipients	% with Pell, 8-digit OPEID match	% with Pell, 6-digit OPEID match
00750100	007501	VATTEROTT COLLEGE	NE	PROPRIETARY	677	68%	68%
02069300	020693	VATTEROTT COLLEGE	IL	PROPRIETARY	1111	215%	76%
02599700	025997	VATTEROTT COLLEGE	MO	PROPRIETARY	5412	371%	73%
02609200	026092	VATTEROTT COLLEGE	IA	PROPRIETARY	1104	271%	69%

IPEDS HD and EFFY Files, 2007-08

8-digit OPEID	6-digit OPEID	UNITID	Institution name	City	State	Sector	12-month undergraduate enrollment
00750100	007501	181756	Vatterott College-Spring Valley Campus	Omaha	NE	Private for-profit, 2-year	991
02069300	020693	148140	Vatterott College	Quincy	IL	Private for-profit, 2-year	516
02069301	020693	437060	Vatterott College	Oklahoma City	OK	Private for-profit, 4-year or above	631
020693A1	020693	450553	Vatterott Education Center	Dallas	TX	Private for-profit, less-than 2-year	309
8-digit OPEID	6-digit OPEID	UNITID	Institution name	City	State	Sector	12-month undergraduate enrollment
02599700	025997	245342	Vatterott College	Berkeley	MO	Private for-profit, 4-year or above	1459
02599701	025997	404383	Vatterott College	Kansas City	MO	Private for-profit, 2-year	1220
02599702	025997	404365	Vatterott College	Springfield	MO	Private for-profit, 2-year	711
02599703	025997	404374	Vatterott College	Joplin	MO	Private for-profit, 2-year	538
02599705	025997	436191	Vatterott College	Sunset Hills	MO	Private for-profit, 4-year or above	977
02599706	025997	440882	Vatterott College-Tulsa	Tulsa	OK	Private for-profit, 2-year	256
02599709	025997	440891	Vatterott College-Wichita	Wichita	KS	Private for-profit, 2-year	370
02599710	025997	440873	Vatterott College-Memphis	Memphis	TN	Private for-profit, 2-year	868

02599711	025997	442408	Vatterott College-Cleveland	Broadview Heights	OH	Private for-profit, 2-year	461
025997A1	025997	445559	Vatterot College-O'Fallon Campus	O'Fallon	MO	Private for-profit, 2-year	531
02609200	026092	373058	Vatterott College	Des Moines	IA	Private for-profit, 2-year	408
02609202	026092	436182	Vatterott College	Saint Joseph	MO	Private for-profit, 2-year	406
02609203	026092	445726	Lecole Culinaire	Saint Louis	MO	Private for-profit, 2-year	777

In cases like this, there is no single variable in the FSA or IPEDS data files which clearly indicates which entries should be grouped together as one school. However, the name, web address, and employer identification number (EIN) is the same for all or almost of these campuses in IPEDS, suggesting that they are all part of one school. The implication of these groupings is that *FSA data such as cohort default rates and loan volume may represent neither an individual campus nor the entire school, but an intermediate grouping* that appears to be based on the historical assignment of OPEIDs.

Campus Groupings Across 6-digit OPEIDs

In some cases, multiple campuses (with unique 6-digit OPEIDs) may report data as a group under one campus' 6-digit OPEID. One such case is the University of Michigan's FISAP data. The FISAP is the application schools file to participate in federal campus-based aid programs. The three University of Michigan campuses file a single FISAP under the University of Michigan-Ann Arbor's OPEID, as shown in the following data file excerpts:

OPE FISAP File, 2007-08

OPEID	Name	State	Zip code	School type	12-month undergraduate enrollment
00232500	University of Michigan - Ann Arbor	MI	48109-1340	Public 4 year	41,833

IPEDS HD/EFFY Files, 2007-08

8-digit OPEID	6-digit OPEID	UNITID	Institution name	State	Zip code	Sector	12-month undergraduate enrollment
00232500	002325	170976	University of Michigan-Ann Arbor	MI	48109	Public, 4-year or above	26,964
00232600	002326	171137	University of Michigan-Dearborn	MI	48128-1491	Public, 4-year or above	7,875
00232700	002327	171146	University of Michigan-Flint	MI	48502-1950	Public, 4-year or above	7,046

There may be other cases like the University of Michigan, but the current configuration of federal financial aid data files make it difficult to determine which matches are appropriate. *The 12-month undergraduate enrollment figure reported on the FISAP exceeds the equivalent figure from IPEDS in many cases when using a 6-digit OPEID match, even when taking into account groupings of campuses in IPEDS with the same 6-digit OPEID. There are also many campuses in IPEDS that have 6-digit OPEIDs that do not appear in the FISAP file.* However, it is difficult to determine if this is because they do not participate in campus-based aid programs or if, as with the University of Michigan campuses in Flint and Dearborn, they really should be grouped with other campuses that file FISAP together.

Recommendations for Improving Usability of FSA/OPE and IPEDS data

In recent years, the Department has taken steps toward increased transparency by posting more institution-level data online than ever before. To ensure that these data are fully usable to answer questions posed by consumers, policymakers and researchers, the Department should make its datasets more compatible so that when combined, they provide comprehensive, comparable, and accurate data on all schools and campuses. By better facilitating the matching process between datasets, the Department would substantially improve both consumers' ability to make informed choices about post-secondary education and training, and researchers' and policymakers ability to identify trends and analyze the impact of federal, state, and institutional policies on student outcomes such as enrollment, persistence, and completions.

At minimum, the Department should provide sufficient documentation to allow users to match data from IPEDS with data extracted from FSA/OPE databases and reports

- For each data year (the academic year that a dataset describes), provide a crosswalk mapping:
 - each 6-digit OPEID in FSA/OPE data to the corresponding UNITID(s) and vice versa; and
 - each 8-digit OPEID in FSA/OPE data to the corresponding UNITID(s) and vice versa.
- Include documentation with each dataset detailing which crosswalk to use for that dataset and any caveats or special notes.

- Clearly document any cases where data reported in one file represent more than one campus.
- Only provide the 6-digit OPEID in FSA/OPE files where data are only reported at the 6-digit OPEID level, not the 8-digit OPEID for the main campus.
- Format OPEIDs as text so leading zeroes appear.
- Clearly document the date at which data were extracted or collected and whether the data might be updated in the future.

In addition, we urge the Department to consider these additional steps to reduce the need for such crosswalks and improve the availability, quality and comparability of data:

- Require reporting of financial aid data based on the *attended* campus/location (rather than the main campus for a group of schools) for all FSA, OPE, and NCES data.
- Adopt common identifiers for each school and campus/location to be used across all datasets.
- Adopt common rules and practices for adding, deleting, and changing the identifier for a given campus or school as changes in status, control, or affiliation occur over time.
- Include a consistent sector designation which is based on program level and type of control (e.g., public 4-year) in all datasets.

This page left intentionally blank.

APPENDIX D:

NPEC FINANCIAL AID DATA WORKING GROUP

NPEC is a voluntary organization established by NCES that includes representatives from federal agencies, postsecondary institutions, associations, and other organizations with interest in postsecondary education data collection. The mission of NPEC is to promote the quality, comparability, and utility of postsecondary data and information that support policy development at the federal, state and institution levels. NPEC is responsible for developing a research and development agenda for the Integrated Postsecondary Education Data System (IPEDS). NPEC monitors and reviews commissioned background papers on topics of importance to the IPEDS Technical Review Panels.

Chair

Matthew Reed

The Institute for College Access and Success

NPEC Members Who Served on the Working Group

Victor Borden

Indiana University, Bloomington IN

Bryan Cook

American Council on Education

Tammy Halligan

Career College Association

Gigi Jones

National Association of Student Financial Aid Administrators

Christine Keller

Association of Public and Land-Grant Universities

Lesley McBain

American Association of State Colleges and Universities

Brian Prescott

Western Interstate Commission for Higher Education

Dawn Geronimo Terkla

Tufts University

Additional Working Group Members and Contributors to This Project

Sandy Baum
George Washington University

Nancy Borkow
National Center for Education Statistics

Archie Cubarrubia
National Center for Education Statistics

Rita Grogan
Mission College, Santa Clara

Mark Kantrowitz
FinAid.org and Fastweb.com

Elise Miller
National Center for Education Statistics

Shirley Ort
University of North Carolina, Chapel Hill

Consultants/Contractors

Brenda Albright
Franklin Education Group

Patricia Steele
HigherEd Insight, LLC