National Center for Education Statistics

U.S. Department of Education
Institute of Education Sciences
NCES 2005–160

Feasibility of a Student Unit Record System Within the Integrated Postsecondary Education Data System

Research and Development Report

March 2005

Alisa F. Cunningham
Institute for Higher Education Policy

John Milam
HigherEd.org, Inc.

Cathy Statham
Project Officer
National Center for Education Statistics

U.S. DEPARTMENT OF EDUCATION
Margaret Spellings
Secretary

INSTITUTE OF EDUCATION SCIENCES
Grover J. Whitehurst
Director

NATIONAL CENTER FOR EDUCATION STATISTICS
Grover J. Whitehurst
Acting Commissioner

The National Center for Education Statistics (NCES) is the primary federal entity for collecting, analyzing, and reporting data related to education in the United States and other nations. It fulfills a congressional mandate to collect, collate, analyze, and report full and complete statistics on the condition of education in the United States; conduct and publish reports and specialized analyses of the meaning and significance of such statistics; assist state and local education agencies in improving their statistical systems; and review and report on education activities in foreign countries.

NCES activities are designed to address high priority education data needs; provide consistent, reliable, complete, and accurate indicators of education status and trends; and report timely, useful, and high quality data to the U.S. Department of Education, the Congress, the states, other education policymakers, practitioners, data users, and the general public.

We strive to make our products available in a variety of formats and in language that is appropriate to a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other NCES product or report, we would like to hear from you. Please direct your comments to:

National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
1990 K Street NW
Washington, DC 20006–5651

March 2005

The NCES World Wide Web Home Page is: http://nces.ed.gov
The NCES World Wide Web Electronic Catalog is: http://nces.ed.gov/pubsearch

Suggested Citation
Cunningham, A. F., and Milam, J. (2005). *Feasibility of a Student Unit Record System Within the Integrated Postsecondary Education Data System* (NCES 2005–160). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

For ordering information on this report, write:

U.S. Department of Education

ED Pubs
P.O. Box 1398
Jessup, MD 20794–1398

or call toll free 1–877–4ED–PUBS; or order online at http://www.edpubs.org

Content Contact:
Cathy Statham
(202) 502–7383
cathy.statham@ed.gov

Executive Summary

This report examines the feasibility of implementing a student unit record (UR) system to replace the student–related components of the Integrated Postsecondary Education Data System (IPEDS). The feasibility study was initiated by the National Center for Education Statistics (NCES), a part of the Institute of Education Sciences (IES) within the U.S. Department of Education (ED), in response to growing interest within the postsecondary education community for more accurate measures of net price and graduation rates, especially measures that take into account institutional mission and student mobility. This interest parallels a growing congressional desire to hold postsecondary institutions accountable for student outcomes.

Background

This discussion of the feasibility of a UR system at the federal level is occurring within the context of the development of other UR systems for students attending postsecondary institutions. Unit record systems are maintained by most colleges and universities to track registration for courses, academic performance, degree and certificate completion, financial aid, and other purposes. A number of states began to develop UR systems in the mid–1980s and use UR data for analysis and program evaluation. Today, 39 states have at least one student UR system. A limitation of state UR systems, however, is that most do not include data on students attending private institutions, or students who leave an institution and transfer across state lines.

Many governmental and other organizations also maintain UR systems on specific groups of students. For example, the National Student Loan Data System (NSLDS) within the office of Federal Student Aid (FSA) compiles information on all recipients of federal student loans, including verification of enrollment by academic term. In addition, the National Collegiate Athletic Association (NCAA) collects UR data on 1,800 institutions with Division I, II, or III varsity athletic programs, and about 2,800 colleges and universities currently contract with the National Student Clearinghouse to perform enrollment verification and other services using student UR data uploaded from member institutions.

At IES/NCES, the Integrated Postsecondary Education Data System (IPEDS) is the core postsecondary education data collection program, designed and implemented to meet its mission to report on the condition of postsecondary education in the United States. IPEDS is a single, comprehensive system that encompasses over 10,000 institutions whose primary purpose is to provide postsecondary education (including roughly 6,700 institutions that have Program Participation Agreements with ED for Title IV federal student financial aid programs and are required by statute to report to IPEDS). The IPEDS system collects institution–level

data in the areas of enrollment, program completions, graduation rates, faculty, staff, finances, institutional prices, and student financial aid. The use of aggregate data has some limitations in comparison with UR data, such as the inability to track the academic progress and experiences of individual students, and therefore to study the longitudinal enrollment of different types of students.

Despite its comprehensiveness, the IPEDS system cannot measure many of the evolving trends in postsecondary education that are necessary for sound policy decisions. The current IPEDS framework cannot accurately capture changing enrollment and completions patterns in the postsecondary education sector, especially given increasing numbers of nontraditional students, and cannot describe the prices various types of students face after financial aid is taken into account. To do so, it would be necessary to collect accurate student–level information on persistence systemwide (i.e., regardless of institution and nationwide), multiple enrollment, part–time enrollment, transfer, and attainment. It would also be necessary to collect student–level information on prices and financial aid, in order to calculate net prices that take into account the individual circumstances of each student. By its very nature, a UR system would enable the collection of data that would lead to more accurate estimates of these variables. In addition, a UR system would allow the development of a whole range of new measures, such as net prices for specific groups of students, graduation rates that take into account institutional missions, persistence rates that consider student mobility and a systemwide perspective, measures of enrollment patterns for nontraditional students, and time to degree by field of study.

Goals and Design of the Feasibility Study

In exploring the feasibility of a UR system, the study attempted to investigate whether such a system *could* be constructed technically and effectively, given the knowledge about UR systems already accumulated at the state and institutional levels. In addition, the feasibility study tried to explore whether such a system *should* be developed by the federal government. To do so, the study solicited input on several dimensions, including privacy and confidentiality, institutional burden, coordination, technical issues, and timing.

As part of the feasibility study, three Technical Review Panels (TRPs) were designed to gather feedback and ideas from different perspectives related to the study, and included representatives from the following groups: 1) states, state systems, private systems, and private associations of colleges and universities; 2) institutions, particularly institutional researchers and registrars; and 3) other stakeholders, including the national postsecondary education association community, federal agencies, units within the ED, and vendors such as administrative information system developers. In addition, the contractor developed an architecture and flow of operations for a proposed student UR system, as well as a list of potential data

elements that might be collected under such a system.

In reading this report, it is important to keep in mind that any redesign of IPEDS to develop a UR system would require legislative authorization through amendments to the Higher Education Act (HEA) and funds would have to be appropriated by Congress to implement the system.

Proposed Redesign of IPEDS

If authorized and funded, the proposed UR system would replace the student–related components in the current IPEDS collection—Fall Enrollment, Completions, Student Financial Aid, and Graduate Rates—as well as the price of attendance variables collected in the Institutional Characteristics component. The UR system would be designed to include all of the variables necessary to replace those components and calculate institution–level estimates for the Peer Analysis System (PAS). The collection process for nonstudent–related components in IPEDS would remain the same.

It is difficult to describe exactly what the UR system would look like before the design process is undertaken. Such a process would involve numerous technical review panels and input from campuses, university systems, and state coordinators, particularly from states with UR systems. Generally, the UR collection system would be designed to collect individually identifiable data through files that are submitted electronically by institutions. The files would be used to calculate institutional summary totals for each school, with information about enrollment, completions, graduation rates, financial aid, and price. Four types of files would be submitted.

- Header files: These data provide individually identifiable information such as name, Social Security Number (SSN), date of birth, address, race/ethnicity, and gender that are attached to an individual student's record. These files would be required at least once for every student. New header records would be submitted as needed to document any changes in these key data.

- Enrollment/term files: These data include program information such as number of courses and credits attempted, major field of study, start and end dates, and attendance status. The files would be required three to four times a year, and institutions would be allowed to upload files more frequently if they wished.

- Completions files: These data include information on degree completions and the date of completion. The files would need to be uploaded at least once per year.

- Financial aid files: These data include information on financial aid received from federal, state, and institutional sources. Information on price of attendance would also be included with the financial aid file. These data also would need to be uploaded at least once a year.

In addition, in the first year of an IPEDS UR collection, additional files would need to be submitted in order for NCES to complete the historical calculations that are part of the Graduation Rate Survey. Depending on program length, these could include up to six years of data for key pieces of information.

For each submission of data, the IPEDS keyholder at an institution or coordinating agency would submit data electronically through the IPEDS collection system, similar to the process that exists currently. After submission, NCES would review the data to make sure they are consistent within the file and with prior submissions. Schools would work with the IPEDS Help Desk to match all records, and any that do not match would have to be resolved. The UR data would then be summarized in online institutional reports, which would also be checked for consistency, before the keyholder "locks" or finalizes the submission.

The UR data would then be moved from the collection system to the permanent database storage system. The full UR database would only exist in this permanent storage area, which would not be accessible via the Internet and would be subject to high IES/NCES levels of protection for confidentiality and security. Ultimately, aggregate estimates would be calculated from the full UR database and moved to the PAS, where they would be stored as institution–level data.

Individually identifiable data would remain within the permanent storage system. The only allowable redisclosures of individually identifiable data would have to be specifically authorized in the HEA legislation, including.

- Enrollment verification for the National Student Loan Data System (NSLDS): The UR system would be used to verify enrollment for students who are receiving federal student loans. Currently, this verification is being done either by institutions themselves, or by organizations such as the National Student Clearinghouse.

- Verification of subsequent enrollment to the IPEDS keyholder: The UR system would be used to redisclose individually identifiable data back to the initial keyholders and to state/system coordinators, in order to give something back to institutions. Data on the subsequent enrollment of students who left the first institution in the previous year would be redisclosed to the keyholder, including the institution of subsequent enrollment, date, attendance status, attainment, and date of attainment.[1]

- Record mismatches: During the process of data collection for the UR system, mismatches between data records and other types of edit failures would have to be resolved. This would involve sending individually identifiable information back to the IPEDS keyholder. These

[1] Redisclosure of student information to the original institutions could take place over a longer time period if this was decided by a future design Technical Review Panels and NCES.

types of edit failure resolutions would be essential to the data integrity of the database.

Other uses of the data would not involve the disclosure of individually identifiable student information. For example, while ensuring the confidentiality of the data, NCES could generate aggregate reports for the Office of Postsecondary Education (OPE) using the UR data (e.g., to generate aggregate measures of persistence, transfer, and attainment for various types of federal student aid recipients, such as those attending on a part–time basis). It would also be possible to add new derived variables to the PAS, used by institutional researchers and other analysts. Each of these derived variables would be reviewed for potential disclosure risks prior to their release on the PAS. Such variables could include new definitions of net price; new measures of graduation rates that better take into account the missions of postsecondary institutions and the mobility of students across institutions; new definitions of time to degree, including transfer calculated for various fields of study; variables that describe enrollment by field of study and program length; and completions by field of study.

Challenges to Implementing a UR System

Technically, UR could be done at most institutions in the long term, after investment of time and financial resources. This can be inferred from the fact that 39 states have compiled UR systems in some form; thousands of postsecondary institutions already submit UR data electronically to private organizations; and postsecondary institutions that are Title IV participants are required to upload information on federal aid recipients to the FSA. Nonetheless, in feedback from institutions, states, associations, and other stakeholders, it is clear if a UR system is legislatively authorized, certain concerns must be dealt with and resolved in the design phase of implementation.

Privacy and confidentiality

Concerns have been raised about student privacy and the confidentiality of individually identifiable student data under a federal UR system. ED, IES, and NCES have always taken seriously the importance of safeguarding student data, but a UR system raises questions about students' rights to withhold or control personal information. This is particularly the case for students who do not receive federal student aid. However, these students benefit indirectly[2] from federal student aid funds, which support all programs, and benefit directly from state appropriations at public institutions and the tax–exempt status of private, not–for–profit institutions. Additionally, data on nonaided students are a critical element to compute graduation rates, retention measures, and other indicators. Information on nonaided students would be necessary in order to compare these measures with information on students receiving student aid.

[2] Tuition at these schools is probably lower than it would be if they were not the beneficiaries of tax–exempt status and state appropriations.

In addition to misgivings about student privacy, there are practical, technical concerns about unauthorized access to the data by hackers and identity theft. This is particularly true given the proposal to use SSNs as one of several personal identifiers that are necessary for matching student records. The use of SSNs would be essential to a UR system to accurately link together student information on financial aid, enrollment, and completions, as well as records from various institutions. Enrollment verification for the FSA already includes the use of SSNs as a student identifier. An additional measure of enrollment intensity at the start of each term (such as full– or part–time) would also be collected to satisfy FSA requirements.

Despite these concerns, IES/NCES is well suited to protect the data, given the strict limits of the legislation regarding data confidentiality under which it operates. IES/NCES legislation protects the privacy of individuals, making wrongful disclosure a Class E felony punishable by up to five years in jail and a $250,000 fine. NCES has experience in working with individually identifiable data through its various sample surveys, and has created the structures and procedures necessary to prevent unauthorized disclosure of such data. In fact, there are no cases where individually identifiable data collected by NCES have been wrongfully disclosed by an employee, a contractor, or a restricted licensee, or of cases in which hackers have breached IES/NCES firewalls. If collected, the data would be technologically protected and secure, and would not leave NCES unless allowed by law. Under the Patriot Act, the Attorney General and the Department of Justice could conceivably obtain access to UR data in order to fight terrorism. Students on whom data are held would be able to "opt out" of the redisclosure of subsequent enrollment information.

Institutional burden

The additional burden of a UR system can be divided into two categories: initial implementation and subsequent operations. The burden of initial implementation is expected to be higher than the costs of subsequent operations. A field test would be necessary in order to make sure the system works, to anticipate and address problems that would be encountered, and to develop all necessary features in the system prior to implementation. About 1,200 to 1,500 institutions would be required to participate in the field test and report using both the old and new IPEDS collection system. Although NCES would make every effort to notify selected institutions early, participating institutions would need to make changes in their reporting systems within a relatively short time frame, depending upon the desire of Congress for an implementation schedule.

In the full–scale implementation, many institutions would need to upgrade information technologies and assign staff to comply with new reporting requirements. Staff would need to be trained in the use of these systems and the details of reporting procedures. Some institutions would need to rely on vendors to provide upgrades to existing

software, build their UR extracts, or pay for changes to legacy information systems. These additional activities would likely increase software costs. Obtaining historical GRS files for all cohorts in the first year would present a burden (although these same files are needed now to calculate the GRS locally). The initial burden on small institutions is likely to be relatively high, unless the institutions are part of a larger system or state association.

The additional costs of subsequent operations under a UR reporting system are expected to be lower than the costs of initial implementation. Keyholders would need to coordinate with offices on campus to gather data, run internal checks to make sure data make sense, submit data to NCES several times per year, and work with the IPEDS Help Desk to reconcile record mismatches and discrepancies in data. Some mismatches of records could be difficult to resolve, especially if there are numerous records.

It is very difficult, at the conceptual stage, to make cost estimates with any degree of precision. Costs would differ widely among postsecondary institutions, depending on whether they are in state UR systems, whether they currently upload to organizations such as the National Student Clearinghouse, whether they use local or proprietary administrative information systems, and the extent of their IT and institutional research capability. There would be a decrease in burden after the initial implementation of a UR system, as postsecondary institutions would no longer need to track and maintain records on GRS cohorts for six years or fill out the current IPEDS student–related components.

If a UR system were implemented, it would be important to take into account these various issues during the design phase of implementation so as to minimize institutional burden. There are different ways to offset the cost and burden of a UR system. One funding mechanism, Administrative Cost Allowances (ACAs), is used to help defray the cost of administering federal student aid programs.[3] A similar funding mechanism could be put in place for a UR system.

Technical issues

Technical issues were also raised as a potential challenge to the implementation of a federal UR system. The proposed system would include the creation and maintenance of a database of millions of student records, with new records added every year. In addition, the system would require the uploading of large files from postsecondary institutions to NCES, using multiple forms of security to protect against unauthorized disclosures of data. NCES currently has most of the hardware and software necessary to implement a UR system, including current equipment used in the web–based IPEDS collection as well as servers capable of storing large amounts of student data. One necessary addition would be database storage, to be located offline in a secure

[3] Institutions currently receive over $150 million in Administrative Cost Allowances (ACAs), which is provided to help cover the cost of administration of federal programs such as Pell Grants and campus–based aid.

site and protected by physical and software firewalls.

There would likely be greater technical challenges for postsecondary institutions, with the extent varying among the registrar, institutional research, and financial aid offices, which sometimes utilize different and incompatible information systems. Institutions using both legacy and proprietary student information systems would need to make software conversions or updates. For the smallest schools, an Excel template could be provided to collect data and generate the data file needed for submission. Although the technical issues could present a problem, these schools currently find a way to do uniform reporting for FSA financial aid eligibility and NSLDS loan deferment.

The proposed UR system would also use XML[4] technology for the submission of data files to NCES, although it is likely that ASCII files would be accepted in the early years of implementation. Some postsecondary institutions have already adopted XML and are using it in their exchange of data with other organizations. On the other hand, many institutions do not currently use XML and training would be required on the use of this technology. Nonetheless, the FSA has already mandated that institutions begin submitting data to the office using XML by 2005–06.

Coordination

Coordination of the flow of information presents a multitude of challenges in implementing a UR system. For example, a UR system might not work well within the existing IPEDS structures in some states. Most state UR systems are based on specific census dates. If multiple header and/or enrollment files need to be submitted at different points in time to capture total enrollment, this would involve a change in workload for both institutions and systems. Special TRP meetings should be held during the UR design phase in order to leverage existing UR systems whenever possible in order to meet both federal and state/system requirements and needs. This will prevent unnecessary duplication of effort and reporting, and ensure that any federal UR system maximizes the lessons that have been learned through years of state UR reporting.

Timing

In implementing a UR system, the timing of data collections would have to be addressed. If a UR system were authorized in 2005, a field test would then be administered in 2006–07, followed by full–scale implementation in 2007–08. The project timetable is designed to yield data relatively quickly while avoiding potential problems associated with an expedited timeframe. A phased implementation could also be considered to provide additional time to address problems during

[4] XML is a "markup language," or mechanism for identifying structures within a document or data file. It employs tags to identify data elements, thereby facilitating the seamless exchange of data. In other words, it allows users to describe data and deliver it across a network, through the creation of common records across disparate databases.

implementation. To respond adequately as part of the field test, it might be necessary for institutions to examine the utility of their administrative information systems for the purposes of producing UR extracts and to address some of the burden issues mentioned above such as training and staffing. Early notification for the selected institutions would be crucial for the institution's ability to respond in a timely and accurate fashion. It is possible that NCES could draw the sample of institutions immediately after legislative authorization to allow selected institutions almost a year to prepare.

Since the UR system is based on individually identifiable records, it must comply with the Office of Management and Budget (OMB) requirement for collecting race/ethnicity data with a two–question format. A by–product of the UR system is that schools that have not yet implemented this change will need to do so to meet OMB Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting.

Another important issue is operational—how to time data collection schedules, while minimizing conflicts with other reporting schedules. The proposed UR system likely would collect enrollment records once per term. However, some institutions do not have standard terms; for example, courses may be offered on a rolling basis or on six–week terms. Institutions could choose to upload data more frequently, especially for the purpose of enrollment verification for student loan programs. It would be necessary to find a method of specifying a whole range of flexible term reporting options, perhaps by asking institutions to document all possible term sequences using the IPEDS Institutional Characteristics component.

Degree and certificate completions would likely be collected with only one file per year, although institutions with several commencement periods might wish to submit multiple files over the year. In some cases, awards are recorded months after the relevant students have stopped attending institutions; degree dates then reflect the date the degree was awarded rather than when the degree was finished. In designing the timing of data collections and the periods of reference for the data, it would be useful to align the completions data with the enrollment data necessary to calculate graduation rates so that completions records can be matched to comparable enrollment records.

Student financial aid information also would likely be collected in only one file per year. Data submitted in an academic year would be from the previous year's award cycle. It would be important to time the collection of financial aid data so that it does not conflict with the institution's aid packaging period, which is the busiest time of year for financial aid offices. In addition, the treatment of summer sessions varies by institution, especially regarding whether summer sessions would follow or lead the submission of an annual data file.

All of these timing issues would be addressed during the design phase of UR implementation, should a UR system be authorized. In the proposed UR system,

collection schedules would not need to be on a uniform schedule, but rather could be geared to a schedule that works best for individual institutions. In other words, institutions with different calendars or financial aid packaging schedules could submit data to NCES on different cycles.

Conclusions

As this report has outlined, a central question for a UR system is *"Could it be done?"* Have the information technologies and infrastructures at the campus and state levels matured, could the current IPEDS web–based reporting system be adapted to a UR system, and would there be adequate technical and legal protections in place at IES/NCES? The report has addressed some of the technical and system problems associated with the design and development of a new IPEDS UR system. At the technical level, a UR system could be done at most institutions given time for implementation.

The feasibility study also addressed the "*Should it be done?"* question, providing a framework for the discussion of issues inherent in this question. These issues constellate in several areas of concern—privacy, burden, coordination, technology, and timing—which would need to be addressed and resolved in the design phase of a UR system should policymakers decide to authorize and fund such a system.

Finally, the feasibility study outlined areas of federal interest: better information for informed consumer decisions, including the improved calculation of net prices; and more accurate measures for institutional accountability and program effectiveness, including enrollment, persistence, transfer, and attainment rates by program of study. Policymakers would be able to monitor in real–time federal student aid programs (such as Pell Grants) and variations in aid packaging. The study also has attempted to highlight some potential benefits to institutions, researchers, consumers, and other users of NCES data.

The study did not attempt to address every challenge or make recommendations about how each aspect should be addressed. Nor did the study document specific organizational positions regarding the obstacles a UR system might face. Rather, it provided a framework for policymakers to understand the potential costs and benefits of a UR system as they discuss whether it should be considered.

The central defining question of the feasibility of a UR system in IPEDS is not a "could" question. It is a "should" question, asking whether the federal government should develop a system that is based upon individually identifiable information about enrollment, financial aid, and attainment. This system would, for the first time, give policymakers and consumers much more accurate and comprehensive information about postsecondary education in this country. Some of the benefits of a UR system include the collection of new data that would measure the success rates of students at institutions to which family and federal

student aid monies flow, provide more accurate consumer guidance, and improve federal programs that support those families and students. In addition to benefits, the feasibility study found a number of significant issues that would need to be overcome before a UR system could be implemented, including objections about student privacy, confidentiality of data, new institutional burdens, coordination within and outside of institutions, and timing issues.

Foreword

This report examines the feasibility of implementing a student unit record (UR) system to replace the student–related components of the Integrated Postsecondary Education Data System (IPEDS). These components currently are based on aggregate institution–level data collected through IPEDS. The feasibility study was initiated by the National Center for Education Statistics (NCES), a part of the Institute for Education Sciences (IES) within the U.S. Department of Education, in response to renewed interest within the higher education community for improved data.

The findings in the report are based on several components. Three Technical Review Panels (TRPs) were designed to gather feedback and ideas from different perspectives related to the study, and included representatives from the following groups: 1) states, state systems, private systems, and private associations of colleges and universities; 2) institutions, particularly institutional researchers and registrars; and 3) other stakeholders, including the national postsecondary education association community, federal agencies, units within the U.S. Department of Education, and vendors such as administrative information system developers. In addition, the experiences and architecture of existing UR systems at the state level, other federal agencies, and private organizations were compiled.

Acknowledgments

The authors wish to thank all of those who contributed to this report. We would particularly like to thank our colleagues at the Institute for Higher Education Policy and at HigherEd.org for their feedback and assistance, including Jamie Merisotis at the Institute for Higher Education Policy, and Connie Holohan, Mike Roma, and Andy Simpson at HigherEd.org. We would like to thank Michael Lance of Benchmark Associates and Tod Massa of the State Council of Higher Education for Virginia. We also are grateful for the feedback and support provided by Cathy Statham, Dennis Carroll, and Susan Broyles at the National Center for Education Statistics (NCES) throughout the project.

We appreciate the time and effort of the individuals who participated in the three Technical Review Panels held as part of this study, as well as the numerous individuals who sent personal comments and thoughts about the study.

We also would like to acknowledge the careful review and thoughtful comments provided by the following individuals at various stages of the study: Paula Knepper, Sam Barbett, Marilyn Seastrom, and Jan Plotczyk from NCES; David Bergeron and Dan Madzelan from the Office of Postsecondary Education; and LeRoy Rooker and Frances Moran from the Family Policy Compliance Office (FPCO).

Table of Contents

Page

List of Tables

List of Figures

Chapter 1—Introduction

This report examines the feasibility of implementing a student unit record system to replace the student–related components that currently are based on aggregate institution–level data collected as part of the Integrated Postsecondary Education Data System (IPEDS). The feasibility study was initiated by the National Center for Education Statistics (NCES), a part of the Institute of Education Sciences (IES) within the Department of Education (ED), in response to renewed interest within the higher education community for improved data. The feasibility study was conducted between October and December 2004. This report describes the findings of the feasibility study.

Overview

If a student unit record (UR) system were to be implemented, it would allow the collection of high–quality data for student–related information in IPEDS, especially related to net prices and graduation rates. By virtue of collecting data at the student level, a UR system would lead to more accurate estimates that take into account both nationwide trends happening *across* institutions as well as developments *within* institutions. The current IPEDS framework cannot accurately capture changing enrollment and completions patterns in the postsecondary education sector, especially given increasing numbers of nontraditional students and the mobility of students. It also cannot describe the prices various types of students face after financial aid is taken into account. In addition to producing the same aggregate estimates that are already collected through IPEDS, a UR system would enable a number of additional estimates that would capture new dimensions of postsecondary education. These new measures could better capture the tracking of students across institutions, unduplicated national headcounts, and compute net prices that take into account student characteristics and enrollment patterns.

In exploring the feasibility of a UR system, this study attempted to investigate whether such a system *could* be constructed technically and effectively, given the knowledge about UR systems already in place at the state and institutional levels. In addition, the feasibility study tried to explore whether such a system *should* be developed

by the federal government. To do so, the study solicited input from various sources on several dimensions of the issue, including privacy and confidentiality, institutional burden, coordination, technical issues, and timing.

Three Technical Review Panels (TRPs) were designed to gather feedback and ideas from different perspectives related to the study, and included representatives from the following groups: 1) states, state systems, private systems, and private associations of colleges and universities; 2) institutions, particularly institutional researchers and registrars; and 3) other stakeholders, including the national postsecondary education association community, federal agencies, units within the U.S. Department of Education (ED), and vendors such as administrative information system developers. (See appendix A for agendas and participants.) In addition, the experiences of specific states, private organizations, and other entities that have built or maintained existing UR systems were compiled. A revision of IPEDS would need to consider the effective practices of already existing UR systems and maintain an ongoing dialogue with State Higher Education Executive Officers (SHEEOs), systems, and the states. Also as part of the feasibility study, the contractor (HigherEd.org, Inc.) developed an architecture and flow of operations for a proposed student UR system, as well as a list of potential data elements that might be collected under such a system.

In reading this report, it is important to keep in mind that any redesign of IPEDS to develop a UR system would require authorization through the Higher Education Act (HEA) and appropriation by Congress. This feasibility study was initiated in order to explore whether a UR system could, in fact, be developed, as well as what types of challenges existed to the successful implementation of such a system.

The study did not attempt to address every challenge or make recommendations about how each aspect should be addressed, but rather provided a framework for policymakers to understand the potential costs and benefits of a UR system as they discuss whether it should be considered.

Reasons for Feasibility Study

The feasibility study, reflecting a renewed interest in a UR system at the federal level, is the culmination of several trends in postsecondary education during the 1990s.

- annual price increases at postsecondary institutions that have exceeded increases in inflation indexes such as the Consumer Price Index (CPI);
- policy concerns about the impact of price increases on consumers and on student aid programs;
- a growing congressional interest in holding institutions accountable for outcomes, starting with graduation rates for student athletes and campus crime reporting;
- a demand for better and more timely data to inform policymaking and consumer choices; and
- the desire of many postsecondary institutions for more accurate measures of net price and graduation rates, especially measures that take into account institutional mission and student mobility.

Congress has attempted to address these trends in several reauthorizations of the Higher Education Act. The 1992 HEA Amendments created a "National Commission on College Costs" to study the problem of annual increases in prices at institutions beyond increases in the CPI. Such increases in price were an issue both for consumers and for Congress, which each year faced increased appropriations for federal student aid programs. The commission's report distinguished between cost and price of attendance, recommended more accurate and timely data on costs, prices, and student aid, and looked at the relationship of student aid programs to cost increases. The 1992 Amendments also included "Student Right–to–Know" legislation, mandating graduation rate information for all students. In response, NCES began the Graduation Rates Survey (GRS) component of IPEDS collecting data on graduation rates on first–time, full–time students, within 150 percent of the nominal time to degree or completion.

The 1998 HEA amendments instructed NCES to conduct a "Study of College Costs," which included an analysis of whether student aid programs were themselves a factor in driving up costs and an analysis of net prices, focusing on the relationship of

rising sticker prices and the differential net price paid by students and their families. Net prices reflect the prices paid after financial aid is taken into account. Although there is substantial debate surrounding the issue of which definition of net price is the best to use in examining access or affordability, there is agreement about the fact that it is difficult, if not impossible, to measure net prices on an aggregate level. The net price study (Horn et al. 2002) showed that prices were flat both for low–income students taking grant aid into account and for middle–income students with both grants and loans.

The 1998 HEA Amendments mandated a redesign of IPEDS, making it a significant element in institutional accountability.[5] IPEDS was charged with collecting data on institutional prices, changes in prices over a three–year period, and student aid. NCES was also tasked with making this consumer information readily available online, along with graduation rates. To fulfill this task, NCES created IPEDS College Opportunities on Line (COOL), the Department of Education's provision of information on all Title IV institutions. IPEDS itself became a web–based data collection, to insure more timely data for policymakers and consumers (see further discussion below). All of these changes reflected congressional interest in accountability and having better information on college prices and net prices.

In the context of the current reauthorization of the HEA, price increases, particularly in the public sector, led some in Congress to consider an "Affordability Index" to reign in price increases. Under the proposal, institutions whose prices had increased more than twice the rate of increase in the CPI over a two–year period would be required to meet higher standards of accountability. They would have had to report on why prices had increased and outline steps to reduce the rate of increase to remain eligible for campus–based portions of Title IV of the HEA, where federal student aid programs are authorized. Although the Affordability Index proposal has been dropped, the interest in outcomes measures such as graduation rates has remained.

For example, a recent report by the Education Trust (Carey 2004), *A Matter of Degrees*, as well as a congressional oversight hearing, have focused on institutional persistence and graduation rates and on the limits of the current IPEDS Graduation Rates component in providing accurate information for institutional accountability. In addition,

[5] The mission of and data collected by IPEDS and its predecessor, the Higher Education General Information Survey (HEGIS), have shifted over time. See further discussion below.

the postsecondary education community has shown a renewed interest in better information on graduation rates, which would include new data on nontraditional students who attend part time or otherwise delay their enrollment and have gaps in attendance, as well as on students who attend more than one institution, transfer, or coenroll at multiple institutions. If Congress were to use graduation rates and time to degree as accountability measures for institutions, the consensus is that the GRS in its present form is inadequate. Mission–specific measures would take into consideration the goals of the institution, such as offering two–year, transfer programs; serving part–time, adult learners; or tailoring workforce, noncredit training to the needs of business and industry. Furthermore, the current system treats nongraduates as dropouts, when they may have in fact persisted or completed their educational program at another institution.

More accurate data are necessary for more nuanced policy decisionmaking, toward the goals of improving student performance and informing students and parents about the true costs of college. The American Council of Education (ACE), the Association of State Colleges and Universities (AASCU), and the State Higher Education Executive Officers (SHEEO) sent letters to congressional leaders, asking that IES/NCES conduct a "feasibility study" of a data system, derived from URs, that would provide mission–specific data on enrollment patterns of all students and outcome measures such as institutional persistence, completion rates, and time to degree, along with detailed information on student aid that would make possible accurate calculations of "net price" for students. Taking into account individually tailored financial aid packages—including the packaging of aid from federal student aid programs—would allow the federal government for the first time to assess accurately the relationship of various student aid programs to persistence.

The debate on the feasibility of a UR system at the federal level is occurring within the context of the development of other UR systems for students attending postsecondary institutions. The following section describes some of these efforts, as well as the current IPEDS framework, future data needs, and the context of IES/NCES operations related to the protection of individually identifiable data.

Background

Unit Record Systems in Operation

It is important to distinguish between the two types of data that are available for analysis: (1) summary or aggregate data; and (2) microdata, the raw or unit record (UR) data that are summarized or "rolled up" into aggregate data. For example, an aggregate report may document the number of bachelor's degrees awarded by an institution, where UR data would document the data that go into the report; in this case, individually identifiable information about each degree recipient. A UR system could then document students' demographic, enrollment, attainment, and financial aid information, as individually identifiable records if desired. For each type of data collected, schools would submit one record per student per term per institutional identifier.

For many years, colleges and universities have maintained computerized recordkeeping through the use of administrative information systems. Typically, specialized admissions software is used to monitor student applications and acceptances, in addition to human resources software for hiring and paying employees and registrar software to keep track of course enrollments, grades, and awards/degrees. In order to submit URs to a federal, state, or other data collection, schools must create electronic extracts or snapshots of their recordkeeping data from these different administrative information systems. These extracts are created using special software tools and can include whatever variables are desired. Extracts represent the selected data and records as of a specific point in time when the files are cut.

In order to complete aggregate summaries in reporting to governing boards, state agencies, or other entities, institutions have two basic choices: (1) run a computer program (or query) against the live, administrative information system to produce a summary report; or (2) create an extract of the data needed and use these records to produce a summary report locally. For both choices, schools must then engage in a significant amount of review and clean up of data to ensure that they can be aggregated accurately to reflect the institution at the point in time (or census date).

In cases in which institutions must submit UR data directly rather than in summary reports, schools have two choices about the initial cleanliness of their file submissions. They may either (1) submit the raw data they obtain from their administrative information systems (which are called transaction files because they

represent ongoing transactions or interactions with the live database); or (2) submit files on which they have conducted further editing (in which case they are called analytical files because they have been scrutinized from an analytical perspective). The different approaches are important to recognize because, if submitted, transaction files may be missing some data or include unexpected values, such as invalid Classification of Instructional Programs (CIP) codes. The resulting summary report may be different than expected for the institution as a whole. However, increased work is necessary to create analytical files, resulting in greater burden on institutions.

The distinction between aggregate and UR data is important because there are inherent limitations to using aggregate data, just as there are issues of data integrity for transaction versus analytical files. Aggregate data collected at one level of analysis cannot be used for lower levels of analysis, such as how the data differ among various groups of students or how students move between institutions (unless this is specifically included at both levels). With aggregate data, it is difficult (if not impossible) to examine relationships among variables or to recompile data for different reporting needs (National Postsecondary Education Cooperative 1998).

Many governmental and other organizations maintain UR systems on specific groups of students. For example, in the Department of Education, NCES conducts sample surveys of postsecondary students (such as the National Postsecondary Student Aid Study, [NPSAS]) in which it collects UR information from the institution for each student in the sample. The National Student Loan Data System (NSLDS) within the office of Federal Student Aid (FSA) compiles information on all recipients of federal student loans, including verification of enrollment by term. FSA also has detailed data on all federally aided students, which represent more than half of full–time undergraduates. The Office of Postsecondary Education (OPE) collects student–level information on the recipients of specific program funds, such as GEAR UP, Upward Bound, and Talent Search, for program evaluation.

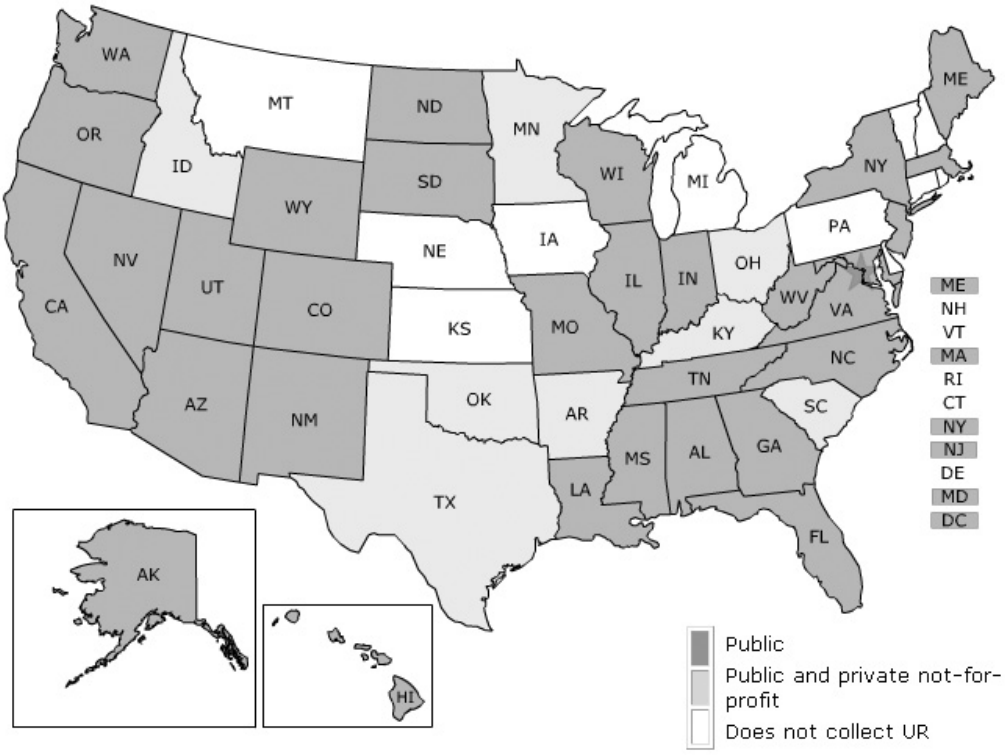
Other branches of the federal government also collect student information. For example, the National Science Foundation (NSF) conducts its Survey of Earned Doctorates annually from all individuals receiving research doctoral degrees from U.S. institutions. The Internal Revenue Service (IRS) requires colleges and universities to annually submit individually identifiable student data on tuition and related expenses and

scholarships/grants, for all enrolled students, in case they or their parents claim a Hope or Lifetime Learning tax credit. This same information also is used to send 1098–T forms to students for use in preparation of their tax forms. More recently, the Department of Homeland Security (DHS) has created the Student and Exchange Visitor Information System (SEVIS) in order to maintain information on nonimmigrant students and exchange visitors from the time they receive their visa documents until they complete their programs. Under SEVIS, colleges are required to provide regular electronic reports confirming enrollment in postsecondary institutions to the DHS.

State and local governments, which are the primary funding sources for public colleges and universities, have tied this funding to requirements for collecting increasing amounts of data on students attending institutions in their states. A growing number of states began to develop UR systems in the mid–1980s, and some states have 30–year histories of using UR systems for analysis and program evaluation. In fact, according to a recent study (Ewell et al. 2003), 39 states have at least one student UR system; some states have more than one, because separate UR systems are maintained for state colleges, community colleges, and other system structures. Most of these states collect data only on students attending public institutions. However, 12 state UR systems include data on students attending at least some private institutions, and the number is growing (figure 1). Most of these state–level UR databases are maintained by state higher education agencies and multicampus postsecondary education systems. In many cases, state UR databases provide extensive support for institutional IPEDS reporting and, in the long run, reduce the reporting burden to individual institutions.

A limitation of state UR systems is that most do not include students attending private institutions, or students who leave and cross state lines. In recent years, there has been substantial interest in the possibility of linking state UR systems to try to minimize these coverage issues. Ewell et al. (2003) noted that in order for existing data in state–level UR databases to be used to track students on a national basis, they must cover a substantial proportion of the nation's enrollment, the systems must contain roughly similar data elements (with similar definitions), and a method must exist to link them consistently. The study found that state–level UR databases consistently track information on enrollment, degree attainment, program, gender, race/ethnicity, and date of birth, and

Figure 1. States with unit record systems



SOURCE: Ewell, P., Schild, P.R., and Paulson, K. (2003). *Following the Mobile Student: Can We Develop the Capacity for a Comprehensive Database to Assess Student Progression?* New Agenda Series. Bloomington, IN: Lumina Foundation for Education.

cover about 73 percent of national headcount enrollment. In addition, definitions and structures are compatible enough to allow linkages among databases. Few of these systems include information on financial aid. The National Center for Higher Education Management Systems (NCHEMS) is currently conducting a follow–up pilot program, funded by the Lumina Foundation, to test the feasibility of voluntary, interstate data sharing that would link 10 state UR databases.

Another experimental study is being funded by the Lumina Foundation to collect UR data from a number of community colleges (JBL Associates 2004). Thirty community colleges are submitting data each term for six student cohorts. Data include student contact information, demographics, transfer credits, current courses, credits, grade point average, student aid, and information on college intervention programs. Lumina and the American Association of Community Colleges (AACC) plan to create a database linking the data together longitudinally and create reports for each participating

college.[6] A similar exchange of UR data by community colleges is being done by Jeffrey Seybert at Johnson County Community College with the League for Innovation in the Community College, as part of the Community College Benchmarking Project. Statewide associations of private colleges and universities in Minnesota and Tennessee have initiated their own collections of UR data recently.

Private organizations also are involved in collecting student–level data. The National Collegiate Athletic Association (NCAA) is experimenting with the use of UR data for student athletes from its member colleges and universities.[7] NCAA collects UR data on 1,800 institutions with Division I, II, or III varsity athletic programs, due in large part to the limits of the current GRS. In addition, many colleges and universities participate in the National Student Clearinghouse, which performs enrollment verification and other services for them using student UR data. The clearinghouse, a nonprofit organization begun in 1993 to facilitate the student record verification process, asks participating colleges to periodically provide them with student UR data electronically. The clearinghouse then provides guaranty agencies and FSA with status and deferment information on student financial aid recipients on behalf of those institutions. The clearinghouse currently is used by about 2,800 colleges and universities (out of the 6,700 Title IV institutions in IPEDS). Many institutions also pay the clearinghouse for information about students' enrollment and attainment after they leave the requesting institutions.

The National Postsecondary Education Cooperative (NPEC) has conducted several studies of UR systems, including the final report of its Working Group on Unit Record Data Versus Aggregate Data (National Postsecondary Education Cooperative 1998), which compared aggregate and UR approaches to data collection. In 2001, another working group examined student transition data systems, including exchange of records across postsecondary institutions (National Postsecondary Education Cooperative 2001). Work on K–16 UR systems was analyzed as part of a recent publication by the State Higher Education Executive Officers (Voorhees and Barnes 2003), which reported that exemplary state data and accountability systems help establish standards, track the

[6] With the assistance of JBL Associates.

[7] NCAA used the IPEDS Graduation Rates Survey figures for its annual collection of graduation rate data between 1998 and 2003. See the NCAA website for more information: http://www2.ncaa.org/.

performance of individuals throughout their educational careers, and increase stakeholders' commitment to gathering and using data on student performance.

It is important for NCES to utilize the extensive experience of the states and others, as well as the known benefits of UR systems, to improve upon its data collection capacity, as part of its ongoing process to improve the analytic capability of IPEDS.

Current IPEDS Framework

NCES has been charged by Congress to report on the condition of postsecondary education in the United States, including changes in its size, participants, providers, and other characteristics. To do this, NCES established the IPEDS as its core postsecondary education data collection program in 1986.[8] Prior to IPEDS some of the same information was collected by the Higher Education General Information Survey (HEGIS). Between 1966 and 1985, HEGIS collected information from higher education institutions that offered courses creditable toward a bachelor's degree. IPEDS is a single, comprehensive system that encompasses over 10,000 institutions whose primary purpose is to provide postsecondary education, thereby expanding the universe of institutions. IPEDS includes about 6,700 institutions that have Program Participation Agreements with ED for Title IV federal student financial aid programs, for which reporting is mandatory.

The IPEDS system is built around a series of interrelated components that collect institution–level data in the areas of enrollment, program completions, graduation rates, faculty, staff, finances, institutional prices, and student financial aid. For example, researchers can use IPEDS to analyze information on enrollments of first–time freshmen by race/ethnicity and gender; institutional revenue patterns by source of income; degrees and certificates awarded by type of program, level of award, and race/ethnicity and gender of recipient; and the characteristics of postsecondary institutions, including tuition, room and board charges, calendar system, accreditation, and price of attendance. Racial and ethnic data collected through IPEDS are used as part of each postsecondary institution's Compliance Report for the Office of Civil Rights (OCR).

[8] Most of the description of IPEDS was taken from U.S. Department of Education, National Center for Education Statistics 2004. More detail can be found at the IPEDS homepage: http://nces.ed.gov/ipeds/AboutIPEDS.asp.

The IPEDS survey data collection occurs at three points during the year (fall, winter, and spring), involving multiple web–based components. Several of these components are student related, in that they request information from the institutions about the activities or characteristics of their students. These student–related components include the following:

- Enrollment (EF). This component collects data annually on the number of full– and part–time students enrolled in the fall in postsecondary institutions in the United States and the other jurisdictions by level (undergraduate, graduate, first–professional) and by race/ethnicity and gender of student. The component has requested the total number of undergraduates in the entering class (including first–time, transfer, and nondegree students) since 2001, to form the basis of the Graduation Rates data. Institutions began reporting first–year retention rates for undergraduate students by attendance status in fall 2003. Racial/ethnic data are collected for the OCR as part of the institution's Compliance Report. Age distributions are collected in odd–numbered years by student level. Data on state of residence of first–time freshmen (first–time, first–year students) and the number that graduated from high school in the past 12 months are collected in even–numbered years. Four–year institutions are also required in even–numbered years to provide fall enrollment data by level, race/ethnicity, and gender for nine selected fields of study—Education; Engineering; Law; Biological and Biomedical Sciences; Mathematics; Physical Sciences; Dentistry; Medicine; and Business Management, Marketing and Related Support Services. Finally, the enrollment component collects the 12–month unduplicated headcount and instructional activity data, which are needed to compute a standardized full–time–equivalent (FTE) enrollment statistic.

- Completions (C). This component collects data annually on recognized degree completions in postsecondary education programs by level (associate's, bachelor's, master's, doctor's, and first–professional) and on other formal awards by length of program. These data are collected by race/ethnicity and gender of recipient and by field of study, which is

identified by six–digit Classification of Instructional Programs (CIP) codes (National Center for Education Statistics 2002b). Completions data on multiple majors are collected by CIP code, award level, race/ethnicity, and gender from those schools that award degrees with multiple majors. Racial/ethnic data on completers are collected in odd–numbered years for the OCR as part of their biennial Compliance Report.

- Student Financial Aid (SFA). This component collects the number of full–time, first–time, degree/certificate–seeking students receiving aid compared to the total number of full–time, first–time students, as well as the number of students receiving each type of financial assistance and the average amount received by type. The types of aid included are federal grants, state and local government grants, institutional grants, and loans from any source.

- Graduation Rates (GRS). The Student Right–to–Know Act of 1990 (SRK) requires Title IV institutions to disclose information about graduation or completion rates to current and prospective students. This component collects data on the number of students entering the institution as full–time, first–time, degree/certificate–seeking students in a particular year (cohort), by race/ethnicity and gender; and the number completing their programs within 150 percent of nominal time to completion. In the GRS, if an institution has a transfer mission, transfer–out students in the cohort should be reported, if the transfers to other Title IV eligible institutions are known to the original institution.[9] The GRS also collects data on the number of students receiving athletically–related student aid in the cohort, and the number of these students completing within 150 percent of nominal time to completion.

- Prices section of Institutional Characteristics (IC). The core of the IPEDS system is the annual Institutional Characteristics component, which acts as the control file for the entire IPEDS system and constitutes the sampling frame for all other NCES surveys of postsecondary institutions. This

[9] For these institutions, known transfer students are considered completions.

component collects the basic institutional data that are necessary to sort and analyze not only the IC database, but also all other IPEDS databases. One section of the IC component collects information on tuition and required fees, room and board charges, books and supplies and other expenses charged to various types of students.

There are also IPEDS components collecting information on institutional revenues and expenditures (Finance) as well as faculty and other staff (Salaries, Fall Staff, and Employees by Assigned Position).

IPEDS data collection is conducted using a web–based data collection system. Each postsecondary institution designates a keyholder, who is responsible for ensuring that data for the institution are submitted as well as editing and "locking" the data.[10] Many states or systems also have one or more coordinators who are responsible for reviewing the data for a specific group of institutions and applying subsequent locks.[11] Together, keyholders and coordinators are referred to as the IPEDS coordination tree.

The process of IPEDS data collection requires a number of steps. Keyholders first enter data through online data entry screens for each IPEDS component, which are tailored to each institution based on characteristics such as degree–granting status, control, and length of longest program offered. In many cases, data from previous years are preloaded on the customized screens for easy reference and comparison purposes. Once the current–year data are entered, the keyholders are required to run edit checks and resolve all errors before locking their data.[12] Once data are locked, they are considered "submitted" and IPEDS Help Desk staff conduct a final review. If any additional problems are detected, the Help Desk staff contact the institutions to resolve any remaining questions. Once data are reviewed and problems resolved, the data are moved from the data collection system to the Peer Analysis System (PAS), which is available on the IPEDS website.[13] At the collection level of the PAS, estimates are available to the

[10] Locking occurs when a keyholder has decided that the data are ready to submit to IPEDS; once locked, the data become read–only and the keyholder no longer can make changes.
[11] Coordinators may provide different levels of review. For example, some may only view data provided by their institutions, while others may upload, review, and lock data.
[12] Edit checks are built into the web–based instrument to detect major reporting errors. The system automatically generates percentages and totals on each collection component. Edit checks compare current responses to previously reported data and look for consistency within and between components. As edit checks are executed, survey respondents are allowed to correct any errors detected by the system.
[13] See: http://nces.ed.gov/ipedspas/.

keyholder; once the estimates are moved to institutional level, they become available to other responding institutions for comparison purposes. After the data have been adjudicated, estimates are moved to the guest level of the PAS, where they are publicly available; these are the "official" estimates.

Creation of the PAS was prompted by the 1998 HEA Reauthorization, which directed the Commissioner of Education Statistics to provide students and families with better information about college costs and prices. In response, the NCES Task Force for IPEDS Redesign recommended the move from paper forms to the web–based collection system (Peng et al. 1999). The task force also recommended the development of a peer analysis system, the provision of help desk support for responding institutions, and specific changes in data content, such as collection of information on financial aid and total price of attendance for first time, full–time degree/certificate–seeking undergraduates. These changes, as well as the suggestion to adopt a process of continuous improvement for the IPEDS system, added accountability and the provisions of information to students and families to the mission of IPEDS.

Other recommendations of the task force were adopted and implemented in subsequent years. NCES developed a searchable website, College Opportunities On–Line (COOL), to provide up–to–date statistics on a broad range of postsecondary institutions for easy access by consumers. The site presents general information about the institution and its mission, as well as data on institution prices, financial aid, enrollment, degrees and awards conferred, accreditation, and types of programs that are offered by the institution. COOL is designed to help college students, future students, and their parents understand the differences among colleges and how much it costs to attend college. The site also provides links to each institution's website, campus crime statistics, and other postsecondary education websites.[14] In early 2004, graduation rates were added to COOL, despite the fact that many colleges perceived them to be an inadequate measure of student outcomes.

The IPEDS Data Analysis System (DAS) was recently released by NCES. This online tool allows users to dynamically generate summary tables for one year of IPEDS data. These summary tables provide sums, counts, and percentage estimates. The DAS

[14] The site is available at http://nces.ed.gov/ipeds/cool/.

allows users to select and regroup categorical variables to produce estimates and to identify ranges of values to form subgroups and estimates.

Complementing these tools is an ongoing process of revision and improvement to the IPEDS collection system, as well as the variables that are calculated for the PAS. To effect change within the IPEDS framework, Technical Review Panels (TRPs) are held on a specific topic to solicit input from postsecondary institutions, associations, researchers, and other stakeholders. The findings of the TRP are summarized and then posted to the NCES website for public comment.

Future Data Needs

Despite the comprehensiveness of the IPEDS system, there are several important items that cannot be calculated accurately from the current data, including net prices, graduation rate measures, and other such variables. The current IPEDS framework uses institution–level aggregates for purposes that would be better served by student–level, longitudinal data collection. These techniques are already being employed by other government and nongovernment organizations to look at student behavior using a secure data collection mechanism. A federal UR system could make several improvements to the data available from IPEDS.

The calculation of accurate net prices has become more important as published tuition and fees ("sticker" prices) have become increasingly unrelated to the price students actually pay to attend college (Horn et al. 2002; Cunningham and Merisotis forthcoming). Currently, sample surveys such as NPSAS can investigate net prices with many different definitions (Horn et al. 2002). Other studies have attempted to measure net price on an aggregate level (College Board 2004; Cunningham and Merisotis forthcoming), but aggregate figures mask the wide variation in net prices paid by students with varying characteristics at different types of institutions. Although a UR system would focus on students already enrolled in colleges, it would improve the calculation of net prices for different types of students annually from a dataset that includes information on all types of financial aid. This would allow the federal government, among others, to more accurately measure the relationship of student aid programs to persistence and other outcome indicators.

The calculation of outcome measures such as graduation rates, persistence, time to degree, and rates of transfer has also become increasingly important, and suffers from similar problems with data collection. As mentioned above, the current IPEDS framework includes a Graduation Rates component that requests graduation rate data for a specific cohort of students—full–time, first–time, degree/certificate–seeking undergraduates enrolled in a specific year.[15] Student data are reported by gender and race/ethnicity, and are separated between those entering bachelor's degree programs and other programs. Institutions then report the number of students in the cohort who complete a degree/certificate within 150 percent of the nominal time to award (for bachelor's degrees, this is usually within six years)[16] These graduation rates are calculated so as to comply with the Student Right to Know Act.

The graduation rates calculated from these data allow researchers to make comparisons between institutions, and within institutions, with regard to gender and race/ethnicity. However, these measures are primarily rates of institutional retention, and largely do not capture what happens to students who leave an institution and may go on to attend another institution and attain a degree. The existing GRS report only counts student completers and transfers in a very narrow fashion. It does not include accurate information about students who stop out, drop out, graduate at a later date, or transfer out unbeknownst to the school. These more complex persistence and attainment data can only be documented with a UR–type system that looks longitudinally across schools to analyze all possible enrollment behavior. GRS graduation rates also measure the educational path of the minority of students who are traditional full–time, degree–seeking students. Yet research has shown that 73 percent of postsecondary students are nontraditional, with characteristics such as part–time attendance and delayed enrollment (Choy 2002). In addition, 40 percent of students now enroll in more than one institution at some point during their progress through postsecondary education, including transfer to other institutions as well as coenrollment (Berkner et al. 2002). These nontraditional paths—along with the increasing price of college—may have encouraged students to

[15] Institutions that predominantly use standard academic terms use a fall cohort, whereas other institutions use a full–year cohort.

[16] Two–year institutions may count as graduates those students who complete the "equivalent of an associate's degree" (i.e., students who complete, within three years, a two–year transfer preparatory program that is acceptable for full credit toward a bachelor's degree and that qualifies a student for admission into the third year of a bachelor's degree program) (Association for Institutional Research 2000).

work while attending college and may have contributed to longer times to degree. These issues are especially relevant for colleges that serve large proportions of nontraditional students, such as public 2–year institutions. Therefore, the current IPEDS framework cannot measure all of the aspects of postsecondary education persistence, attainment, and transfer that are necessary for sound policy decisions. To do so, it would be necessary to collect accurate student–level information on persistence systemwide (i.e., regardless of institution, nationwide), multiple enrollment, part–time enrollment, transfer, and attainment, on an annual basis.[17]

Additional information on net price, graduation rates, transfer rates, time to degree, and enrollment by discipline would provide consumers, policymakers, and others with more complete information on which to base decisions. By its very nature, a UR system would enable the collection of data that would lead to more accurate estimates of variables currently in the PAS, such as fall enrollment and institutional graduation rates. Like any data collection instrument, a new UR system would have to recognize some degree of error in measurement; however, a UR system would be expected to produce estimates superior to the ones created under the current IPEDS framework, which, for example, do not capture student transfer in the calculation of graduation rates. Perhaps more important, a UR system would allow the development of a whole range of new measures that could be added to the PAS, such as net prices for specific groups of students, graduation rates that take into account institutional missions, persistence rates that consider student mobility and a systemwide perspective, measures of enrollment patterns for nontraditional students, time to degree by field of study, and other critical data. A variety of factors impact net prices and graduation rates, many of which would not be collected by the proposed UR system. However, the level of detail collected through a UR system would greatly improve understanding of the issues.

There are several other important improvements that would be made by a national UR system. For example, the UR system would expand upon the current Enrollment component to allow unduplicated 12–month enrollment calculations and measures of enrollment by field of study, while continuing to comply with Office for Civil Rights requirements. The current Enrollment component of IPEDS double counts students who

[17] Note that a change in the required calculation of graduation rates would require a change in the Student Right–to–Know legislation.

are coenrolled, and does not provide enrollment counts for all programs (enrollment is requested in a limited number of programs in even–numbered years). Better estimates of full–time–equivalent enrollment (FTE) would also be possible. A national UR system would expand upon the current Completions component by enabling the measurement of time to degree for each field of study. It also would facilitate the calculation of completions for students who attained awards at more than one institution, or attended institutions in more than one sector.

Context of IES/NCES Operations

The creation of a national UR system would involve the collection of individually identifiable student data on a scale that has not yet been accomplished. IES and NCES are required by law to protect the confidentiality of these individual respondents. The collection, reporting, and publication of data based on student records are overseen by the Family Educational Rights and Privacy Act (FERPA). FERPA was enacted in 1974 and is administered by the Family Policy Compliance Office (FPCO) of the U.S. Department of Education (ED). FERPA requires the protection of student records by educational institutions that receive funds from ED, and prohibits the disclosure of individually identifiable information from educational records without students' consent (or their parents or guardians). There are exceptions in specific circumstances, including the disclosure of directory information,[18] disclosure to school officials with legitimate educational interests, and disclosure to organizations conducting studies on behalf of educational agencies or institutions.[19] In addition, FERPA law was amended by the U.S. Patriot Act of 2001, so that any data protected by FERPA are available to the Attorney General for the purpose of investigating or prosecuting acts of terrorism.

In addition, IES and NCES operate under a number of other laws and regulations governing the confidentiality and security of individually identifiable data. The Privacy Act of 1974, as amended, requires federal agencies to collect, maintain, use, or

[18] Directory information includes information that would not be considered harmful if it were released, such as name, address, telephone number, date of birth, field of study, and degrees awarded (Pfeiffer 2003). Also see http://www.ed.gov/policy/gen/guid/fpco/ferpa/index.html.

[19] Under the proposed UR system, the original submission of student records by schools to NCES would be considered disclosures allowable under FERPA. Redisclosures refer to cases in which UR data that have been submitted to the UR system are sent, or "redisclosed," to the original institution or to another party. See chapters 2 and 4 for a more complete explanation.

disseminate any record of identifiable personal information only for necessary and lawful purposes and with adequate safeguards to prevent the misuse of data. A federal agency cannot disclose individually identifiable information without the prior written consent of the individual, although there are exceptions. The Federal Statistical Confidentiality Order of 1997, an order by the Office of Management and Budget (OMB), defines relevant terms and provides guidance on the content of confidentiality pledges that federal statistical programs should use under two different conditions—when the data may only be used for statistical purposes, and when the data are collected for statistical purposes but the agency is compelled by law to disclose the data. The second condition is relevant under the Patriot Act, which permits the Attorney General to petition a court of competent jurisdiction for an ex parte order requiring the Secretary of the Department of Education to provide data relevant to an authorized investigation or prosecution of an offense concerning national or international terrorism.[20]

NCES has had strong confidentiality laws since 1988, where disclosure or publication of individually identifiable information is a Class E felony. The Patriot Act amended the National Education Statistical Act of 1994, which was adopted and amended in the creation of the National Education Sciences Act of 2002. Under the Education Sciences Reform Act of 2002 (ESRA 2002), all individually identifiable information about students, their families, and their schools shall remain confidential. This law requires that no person may use any individually identifiable information for any purpose other than the statistical purposes for which it is supplied. In addition, the law prohibits publications where the data provided by a particular person could be identified, and forbids anyone other than the individuals authorized by the Director of the Institution of Education Sciences to examine the individual reports. Employees or other individuals who knowingly disclose or publish any individually identifiable information are subject to fines of up to $250,000, or up to 5 years in prison, or both. Similarly, the E–Government Act of 2002, Title V, Subtitle A, Confidential Information Protection (CIP 2002) declares that all individually identifiable information supplied by individuals or institutions to a federal agency for statistical purposes must be kept confidential and may only be used for statistical purposes; any willful disclosure of the data for

[20] This law was incorporated into Education Sciences Reform Act 2002.

nonstatistical purposes, without the informed consent of the respondent, is a Class E felony.

As a result of these laws, individually identifiable data must be held confidential unless the individual provides written consent, except for specific exceptions (e.g., the release of directory information under FERPA, the use of information for statistical purposes, and the authorized investigation and prosecution of terrorism). To assist IES/NCES staff and data users in meeting the requirements to protect these confidential data, IES/NCES has established statistical standards that govern the handling and use of confidential data (National Center for Education Statistics 2002).[21] For example, all IES/NCES staff and relevant contractors are required to pledge (and sign notarized affidavits) not to release any individually identifiable data for any purpose, to any person not sworn to the preservation of confidentiality. All materials with individually identifiable data must be kept secure at all times through the use of passwords and secure data handling and storage. In order to prevent the publication of individually identifiable data, NCES uses techniques such as range restrictions, collapsing of categories, data swapping, and other methods of data perturbation. Qualified external researchers who desire access to data files with individually identifiable information must pass through a formal application process for a restricted data license, and must follow similar security procedures as mentioned for IES/NCES staff and contractors, subject to unannounced inspections. The only UR data file that would be made available to licensed researchers is the one used to create the NPSAS sample. The complete UR data would not be available to licensed researchers.

NCES has over three decades of experience with individually identifiable data files, including the National Postsecondary Study Aid Study and other sample surveys. These files have undergone disclosure avoidance procedures prior to release, and restricted use licenses have been granted to qualified researchers. To date, IES/NCES has experienced no known disclosures of the confidential data entrusted to IES/NCES by the many respondents to IES/NCES data collections.

[21] NCES recognizes that the use of UR records for FSA enrollment verification is a nonstatistical purpose and therefore is not protected by CIPSEA. However, the rest of the UR data uses are protected in this manner.

Organization of this Report

The remaining sections of this report describe various aspects of a proposed UR system. The following chapter provides a brief depiction of what a UR system might look like, including the potential redisclosures that might be allowed under such a system as well as some proposed alternatives to a UR system. Next, chapter 3 presents a number of potential challenges to a national UR system, including privacy and confidentiality issues, increased burden to postsecondary institutions, technical problems that might be faced, coordination issues, and possible problems with the timing of data collection. Chapter 4 describes a potential architecture for a UR system, including the collection system, the permanent UR storage database, migration of data to the Peer Analysis System (PAS), redisclosures, training, the IPEDS Help Desk, software, and hardware. The final chapter summarizes the findings of the feasibility study.

Several appendixes also are included with this report. Appendix A provides the agendas and lists of participants from all three Technical Review Panels. Appendix B offers some rough estimates of the potential burden of implementing a UR system.

Chapter 2—Proposed Redesign of IPEDS

The UR system, as proposed, would replace the student–related components in the current IPEDS collection—Fall Enrollment, Completions, Prices, Student Financial Aid, Graduate Rates, and the price of attendance component. The UR system would be designed to include all of the variables necessary to replace those components and calculate institutional–level estimates for the IPEDS Peer Analysis System and other required reporting. At the same time, the collection process for nonstudent–related components in IPEDS (Institutional Characteristics, Finance, Fall Staff, Salaries, and Employees by Assigned Position) would remain the same.

It is difficult to describe exactly what the UR system would look like before the design process is undertaken. In general, however, UR collection may be described as a process involving the upload of individual student records to NCES by the IPEDS keyholder[22] at each institution. The following presents a broad overview of the proposed UR system; Chapter 4 describes the system architecture and collection process in more detail. Note that the proposed UR system outlined here and in Chapter 4 is presented to provide a picture of what a UR system could look like, if authorized, and to help the reader understand the challenges to the implementation of such a system as well as potential solutions. If a UR system were authorized and funded, various aspects of the proposed system could change in the design and implementation phase[23].

Brief Description of Unit Record System

General Architecture

The heart of the proposed UR system involves the collection of individually identifiable records about students based on their enrollment, price paid, financial aid, and attainment at different points in time. When combined for all students and all

[22] The IPEDS keyholder is the person at each institution (or coordinating body) who is responsible for compiling data for IPEDS, submitted the data to NCES, and locking the data.
[23] If a UR system were implemented, TRPs would be held to advise on various aspects of the design of the system. These TRPs would include input from financial aid officers, institutional researchers, registrars, IT staff, and other institutional staff representatives.

schools, the UR database would allow NCES to complete its IPEDS enrollment, financial aid, completions, and price survey components for each institution's data in the PAS and on COOL. In order to collect all of these data, different collection schedules would be required with separate file submissions. Details on the proposed collection schedule are included in Chapter 4. The working list of variables includes the following:

- Student identifiers and demographic information: These data provide individually identifiable information such as name, Social Security Number (SSN) or Individual Taxpayer Identification Number (ITIN), date of birth, address, race/ethnicity, and gender that are attached to an individual student's record. Many of these variables allow records to be matched with one another to follow a student over time, or to follow a student that attends more than one institution. A header file containing these variables would be submitted for each student enrolled at each institution.

- Enrollment variables: These data include program information such as number of courses and credits attempted, major field of study, start and end dates, and attendance status. These data would be required three to four times a year (in other words, once per term), while institutions that do not use a term system would be allowed to upload files more frequently if they wished. The use of multiple files is tied to the FSA enrollment verification process and allows for the more accurate calculation of full–year enrollment. Institutions would identify the type of calendar system they use and the way they measure course activity in the Institutional Characteristics component during the fall collection schedule.

- Attainment variables: These data include information on degree completions and the date of completion. The file would need to be uploaded at least once per year.

- Financial aid variables: These data include information on students' annual receipt of financial aid from federal, state, and institutional sources (i.e., financial aid passing through the institution's financial aid office). These data also would need to be uploaded at least once a year.

- Price of attendance variables: Information on annual price of attendance would be uploaded with the financial aid file.

In addition, a one–time only collection of historical information on GRS cohorts would be required in order to complete the calculations for the GRS. Data would be needed for the established GRS cohorts, including basic enrollment and attainment information. The full student records for each cohort would not be necessary, such as field of study codes and financial aid.

For each submission of data, the keyholder at an institution or coordinating agency would upload a data file using XML,[24] although ASCII would likely be accepted in the initial years of implementation.[25] The data would be submitted through the IPEDS collection system, similar to the process that exists currently. At NCES, the data would go through internal and external edit checks, and mismatches would be identified and sent back to the keyholder for review and resolution. For example, if a student changes her name, the record may show up as a mismatch because all other aspects of the previous header file are the same, but the name differs. Other mismatches might include misspelled names, students identified as first–time students who appear to match records of previously enrolled students, and keystroke errors. When edit failures are resolved or signed off by the keyholder, aggregate reports would be available onscreen to the keyholder to view and correct as needed (see the Chapter 4 for a full description of this process). Once the data are satisfactory and pass all edit checks, they would be locked by the keyholder (and other members of the IPEDS coordination tree, if applicable). The data would then be moved from the collection system to the permanent database storage system. The full database would only exist in this permanent storage area, which would not be accessible via the Internet and would be subject to NCES' high levels of protection for confidentiality and security. The aggregate estimates that are calculated from the full UR database would be moved to the Peer Analysis System, where they would be stored and available as institution–level data.

[24] XML is a "markup language," or mechanism for identifying structures within a document or data file. It employs tags to identify data elements, thereby facilitating the seamless exchange of data. In other words, it allows users to describe data and deliver it across a network, through the creation of common records across disparate databases.

[25] Beginning is 2005–06, FSA will require that all Title IV institutions transmit FSA–required data using XML tags. Some institutions have already begun the process of converting to XML tags. An Excel spreadsheet would be provided for institutions that are not capable of cutting extract files; this would include a feature that lets the user automatically export the spreadsheet data in the required format.

Redisclosures

As mentioned, the UR system would be heavily protected from unauthorized abuse of individually identifiable information. According to FERPA, redisclosures refer to cases in which UR data that were submitted to the UR system by institutions would be sent, or "redisclosed," from the UR system to the original institution or to another party. The only allowable redisclosures of individually identifiable data would have to be specifically authorized in the HEA legislation. Several authorized redisclosures of data have been proposed:

- Enrollment verification for the National Student Loan Data System (NSLDS): The UR system would be used to verify enrollment for students who are receiving federal student loans. Currently, this verification is being done either by institutions themselves, or being outsourced to organizations such as the National Student Clearinghouse.

- Verification of subsequent enrollment to the IPEDS keyholder: The UR system would be of significant value to colleges and universities if they could gain information on students who left their institutions—i.e., if they re–enrolled subsequently or enrolled and obtained a degree from another institution. Therefore, a redisclosure of individually identifiable data back to the initial institution keyholder and coordinator is proposed. Data on the subsequent enrollment of students who left the first institution in the previous year would be redisclosed to the keyholder and coordinator, including the institution of subsequent enrollment, date, attendance status, attainment, and date of attainment.[26] This benefit would come in addition to potential new aggregate variables on graduation rates, time to degree, and student mobility that would be posted to the PAS (see below).[27]

- Record mismatches: During the process of data collection for the UR system, mismatches between data records and other types of edit failures would have to be resolved. This would involve sending individually

[26] Students would be able to opt out of this redisclosure. This will be discussed in future sections.
[27] Redisclosure of student information to the original institutions could take place over a longer time period if this was decided by a future design TRP and NCES.

identifiable information back to the IPEDS keyholder. These types of edit failure resolutions would be essential to the data integrity of the database.

No other new redisclosures of data would be allowed besides those described above and those permitted under the Patriot Act. Individually identifiable data would remain within the permanent UR storage system.

Originally, a proposal was considered to redisclose UR data to the Internal Revenue Service (IRS) to verify enrollment for students who claim the Hope or Lifetime Learning tax credits. However, after considering feedback from TRP panelists as well as discussions with IRS, it was decided not to include the collection of data for the 1098–T forms in the proposed UR system framework. Nonetheless, including 1098–T data in the UR system, if authorized, might save money for the federal government and may be considered in the future.

Analysis for OPE

Individually identifiable student information would not leave the UR database for any other purposes, such as generating performance measures for federal financial aid programs; rather, aggregate reports would be generated by NCES from the data. For example, the Office of Postsecondary Education (OPE) is required to assess the success of student financial aid programs under the Government Performance and Results Act (GPRA). Accurate measures of the persistence and attainment of students who receive federal financial aid are important to fulfill these requirements. OMB has developed annual program indicators for each federal agency under the Program Assessment Rating Tool (PART). For OPE, required program indicators include the percentage of aid recipients that persist and attain a degree or certificate, for various groups of students broken down by attendance status, gender, race/ethnicity, and other factors. Currently, OPE can obtain these types of measures only from the Beginning Postsecondary Students (BPS) study, which is administered by NCES once every six to eight years. A UR system, on the other hand, would allow the calculation of these measures on an annual basis. NPSAS and BPS samples would be drawn from the UR data. NPSAS collects financial aid and employment data well beyond the data that would be supplied with the proposed

UR system. BPS includes a number of variables that are used to predict persistence, including information about family formation, health, and employment.

Under the proposed UR system, NCES could not allow OPE access to the UR database. However, NCES could generate aggregate reports for OPE using the UR data. In other words, OPE could submit the identifiers of student financial aid recipients to NCES, or records could be extracted by NCES from OPE databases. NCES could then match those records with the UR database and generate aggregate measures of persistence, transfer, and attainment for various types of aid recipients, such as those attending on a part–time basis, or those attending institutions in different sectors. Under such a scenario, NCES would conduct disclosure risk avoidance analyses of these aggregate measures and perturb the data as necessary to ensure confidentiality.[28] While ensuring that cell sizes were large enough so that no individual could be identified, NCES would send reports with these aggregate measures back to OPE. A similar process could be performed for program evaluation of other OPE programs, such as GEAR UP, Upward Bound, and Talent Search.

Other Possibilities

After the full–scale implementation of a UR system, it would be possible to add new derived variables to the PAS, COOL, and the Data Analysis System (DAS). These new variables could take into account institutional missions, transfers, or the characteristics of various groups of students. The procedure for adding or modifying data elements under IPEDS involves holding Technical Review Panels to discuss a particular topic and posting the findings of the TRP online for public review and comment.

Several variables of interest were mentioned at the three TRPs. For example, panelists were interested in new definitions of net price. A UR system would allow the development of net price calculators that could estimate net price for groups of students with differing characteristics. After a number of years of implementation of a UR system, it would be possible to calculate new measures of graduation rates that better take into

[28] In order to preserve the confidentiality of information regarding individuals' financial aid and graduation rates, the student financial aid and graduation rate data files are perturbed in various ways. All aggregate student financial aid and graduation rate statistics are based on the perturbed data. NCES would keep a copy of the file with flags for perturbed cells for internal documentation purposes.

account the missions of postsecondary institutions and the mobility of students across institutions. New definitions of time to degree, including transfer, calculated separately for various fields of study, could also be developed. In addition, variables could be created to describe enrollment by field of study and program length, completions by field of study, and other factors.

Alternatives using IPEDS Aggregate Components

Rather than redesigning IPEDS to create a UR system, the current IPEDS framework could be modified by adding additional variables or survey components. Additional data elements might lead to improved aggregate measures, although they would still suffer from the limitations of aggregate data.

For example, a Technical Review Panel was held as part of the ongoing IPEDS revision process to discuss the calculation of net price and the potential for improved data collection. The TRP came up with more than 100 different definitions of net price that would be useful to calculate. TRP members also came to the conclusion that in order to calculate additional definitions of net price that were useful to the postsecondary education community, an entirely new IPEDS net price component would be needed. Such a component would require collection and reporting of tuition and fees, housing costs, books and supplies, and total price of attendance by such categories as residency status, dependency status, attendance intensity, and aid status. Although this option has the benefit of fitting within the current parameters of the IPEDS framework, it would require substantial additional burden on the part of colleges and universities.

To calculate graduation rates and variations on persistence, transfer, and time to degree, it is possible that NCES could modify the Graduation Rates component in several ways. For example, aid categories could be added to the existing matrix of race/ethnicity and gender counts. Transfer counts could be made mandatory for all institutions with transfer missions, and the time to degree could be extended to 250 percent of normal time to degree. Cohorts of part–time students could be established. In addition, a variety of derived variables could be defined and added. Like the net price component described above, however, these changes would involve increases in institutional burden and would not answer many of the questions related to student mobility, multiple enrollment, and

related issues. Changes also would be necessary to the Student Right–to–Know legislation to define new measures of postsecondary outcomes. After legislative change, additions would need to be made to the current IPEDS collection schedule to collect the new variables required to calculate these measures.

To gather better data on completions and enrollment, existing IPEDS components could be modified with increases in data elements. For example, fall enrollment data could be collected by field of study, and age and residence and migration data could be collected annually rather than every other year. Unduplicated full–year enrollment, including credit activity, could be collected by gender, race/ethnicity, student level, attendance intensity, and field of study. Completions data could be collected by field of study and the average number of credits earned, including transfer credit. Again, however, these changes in aggregate data collection would involve increased costs. A UR system would allow the collection of all of these data elements within the framework of fewer file submissions and without adverse impact on the schools least able to do aggregate IPEDS reporting.

Other Alternatives to Unit Records

There are other alternatives to a full UR system that might be able to generate variables needed to fill gaps in necessary knowledge. However, each of these alternatives has its own costs and/or disadvantages. For example, one suggestion for gathering more accurate net price information was to administer the National Postsecondary Student Aid Study as an annual sample survey. NPSAS collects information on students' receipt of various forms of financial aid (federal aid as well as state, institutional, and private), tuition and fees, total prices of attendance, residency, demographic information, and other variables. However, while NPSAS could provide nationally representative net price figures, it could not provide net prices by institution. In addition, representation by state would likely be far more expensive than a UR system, and would be more burdensome for state agencies.[29] NPSAS and BPS are necessary, though, to collect additional data

[29] It was estimated that administering NPSAS on an annual basis for all 50 states would cost an additional $40 million per cycle.

that are needed to understand student persistence, including income, health, and employment.

Another alternative that would allow more accurate graduation rates is that NCES could administer the Beginning Postsecondary Students (BPS) longitudinal study on an annual basis.[30] BPS captures a national, systemwide perspective of persistence systemwide (i.e., regardless of institution), multiple enrollment, part–time enrollment, transfer, and attainment using students as the unit of analysis. However, BPS is administered only once every six to eight years. In addition, similar to NPSAS, BPS is a nationally representative sample of students, and the data cannot be used to make generalizations about states or institutions. Finally, the cost of administering BPS is quite high, and conducting such a survey on an annual basis would be substantially more expensive than a UR system.

It also has been suggested that since the National Student Clearinghouse already collects UR data from 2,800 institutions, perhaps these data could be used for the purpose of calculating graduation rates and other measures. Currently, these data cannot be used to calculate national estimates because the coverage is not complete (there are about 6,700 Title IV institutions), with 4–year institutions being more likely to participate than 2–year institutions or private for–profit institutions. The Clearinghouse currently includes only a subset of the data elements that would be required to complete the existing IPEDS components. In addition, not all participating institutions currently report degree or certificate attainment to the Clearinghouse. This means that students who are enrolled at a nonparticipating institution, or whose degree was not reported to the Clearinghouse, would show up as stop–outs or drop–outs from postsecondary education. Near 100 percent participation would be necessary for the Clearinghouse to be used as a viable alternative. There also are questions about liability and whether these data are protected by FERPA.

[30] BPS is a subset of the National Postsecondary Student Aid Study (NPSAS).

Chapter 3—Issues Related to the Development of a Unit Record System

As mentioned in the introduction to this report, it was important to examine two separate questions in examining the feasibility of a UR system. First, *could* a UR system be technically and efficiently constructed, or are there insurmountable issues to developing such a system? Second, *should* such a system be developed by the federal government? If the barriers to development of the system can be overcome, what are the specific issues that must be addressed before moving forward?

The answer to the first question may be inferred from the fact that 39 states have compiled UR systems in some form, and thousands of postsecondary institutions submit UR data electronically to private organizations such as the National Student Clearinghouse, which collects student UR data for a large proportion of currently enrolled students (in addition, for a fee the Clearinghouse will send back to the institution selected information on their students). Technically, UR could be done at most institutions in the long term, after investment of time and financial resources. NCES already collects student UR data through sample surveys such as NPSAS, and postsecondary institutions have experience in reporting data on financial aid recipients to FSA.

The answer to the second question is not as clear. This chapter provides a framework for the problems associated with the "should" question, including several dimensions: privacy and confidentiality; institutional burden; coordination; technical issues; and timing. Much of the information in this chapter came out of the three TRPs convened as part of the feasibility study, as well as individual comments submitted to the contractor in response to the posting of public documents. Some of the concerns raised in this chapter are addressed more fully in the following chapter, which presents a detailed architecture of a proposed UR system.

One should note that several design issues were raised during the TRPs, often regarding the definition of data elements. These types of questions would be resolved in

the design phase of a UR system, should a UR system be authorized and funds be appropriated.

Privacy and Confidentiality

Some of the biggest challenges to a national UR system are concerns about student privacy and the confidentiality of individually identifiable student data. Some panelists at all three of the TRPs raised the issue of the individual right to withhold or control personal information. If NCES collects individually identifiable data in a UR database, student data would be in the possession of an external party, but may still be within the student's right to control. ED, IES, and NCES have always taken seriously the importance of safeguarding student data, but many in the postsecondary education community are concerned that creation of a federal UR database of all students would potentially be more dangerous than smaller databases held by states or other organizations. Some critics of a federal UR system believe that the simple existence of such a database is a violation of privacy. As one TRP member asked, does the need for data outweigh individual freedoms? Another way to look at this issue is, if a UR system is implemented, can student privacy and the confidentiality of student records be protected?

The goal of the proposal is not to build a system that would endanger students' privacy, but rather to use the experiences of partial collections of student UR data to construct a secure system for collecting data on all students. Currently, information about federal financial aid recipients, including their SSNs, is collected by FSA to evaluate and monitor federal student aid programs. Students who apply for federal student aid or claim the Hope and Lifetime Learning tax credits in effect give consent through the Free Application for Student Aid (FAFSA) and tax forms submitted. Some panelists, however, questioned the need to report data on students who do not receive federal student aid, asking what the compelling government interest is in collecting data on nonaided students and wondering whether the involuntary inclusion of such students violates their rights of refusal. Nonetheless, data on nonaided students would be necessary to compute graduation rates, retention measures, and other indicators in order to compare these measures to those of aided students. An additional argument made was that students who

currently do not receive federal student aid are aided indirectly through such forms of subsidy as state appropriations to public institutions and deferred tax revenues at private, not–for–profit institutions.[31]

Another concern raised regarded the redisclosure of individually identifiable information for purposes of matching student records and providing information about subsequent enrollment of students to IPEDS keyholders. Such redisclosures could provide new information going to the original institutions, for which the students did not specify their consent.

Federal privacy laws, such as FERPA, require that individually identifiable student information collected by federal agencies be protected and released only with the prior consent of the individuals, with certain exceptions. Under FERPA and other privacy laws, organizations are required to notify students if their information may be collected and used for research studies or other purposes. Students would be able to "opt out" of the redisclosure of subsequent enrollment information back to the original keyholder.[32]

Finally, some postsecondary leaders for private institutions are concerned about possible legal liability once they have submitted data to NCES if the data are subsequently misused or unlawfully disclosed, because they do not have the sovereign immunity protection that exists in the public sector. The Office of General Counsel at the Department of Education has agreed that because NCES adds value to submitted data, NCES therefore "owns" the data submitted by the institution.[33] If a UR system were to be legislated and implemented, IES/NCES would work with the FPCO to ensure that collection of data and redisclosures are lawful under FERPA.

TRP panelists also expressed concern over possible unintended uses or consequences of the UR data. Other federal agencies might want access to the data for noneducation–related purposes. This concern is understandable, although it is the purview of Congress to determine legitimate disclosures. If other agencies were to have access to the data, this would need to be incorporated into law as appropriate.

---

[31] Tuition at these schools is probably lower than it would be if they were not the beneficiaries of tax–exempt status and state appropriations.

[32] An "opt out" flag would be included in each student's header record. Students would not be able to opt out of the statistical use of the data. Note that students receiving federal loans have already given their consent to enrollment verification. See the Chapter 4 for details.

[33] This agreement was obtained in relation to NPSAS data from the University of Michigan in 2004.

In addition to misgivings about student privacy, there are more technical concerns about unauthorized access to the data by hackers and identity theft. This is particularly true given the proposal to use SSN/ITINs as one of several personal identifiers to match student records. Many institutions do not require that students provide SSN/ITINs upon admission unless they apply for federal student aid or the Hope and Lifelong Learning tax credits. In fact, several panelists in the second TRP noted that their institutions are moving away from using SSNs to identify student records due to concerns about unauthorized access, and that more students are refusing to supply those numbers. TRP panelists suggested looking into alternatives to collecting SSNs, such as using compression or only the last six digits of the number. Nonetheless, SSN/ITINs are currently required by FSA to determine aid eligibility and by NSLDS for loan deferment. The use of SSN/ITINs would be essential to a UR system, to accurately link together student information on financial aid, enrollment and completions, as well as to link records from various institutions. Without SSN/ITINs, mismatch rates would increase significantly, which would substantially increase the burden on the Help Desk and keyholders to resolve. The practice of obtaining SSN/ITINs from students would have to be reinstated if the UR system were authorized. For students without SSN/ITINs or who refuse to provide them, a matching process would be used with fuzzy logic and identifiers such as name, date of birth, and address. Fuzzy logic involves mathematical algorithms or methods for making a decision (in this case a fuzzy match between records) based on ambiguous or missing information.

Certainly there are valid concerns about privacy in a time when increasing amounts of information are being gathered on all citizens as well as students. Nonetheless, there are assurances IES/NCES can make regarding the confidentiality of any data collected through a UR system. IES/NCES is well suited to protect the data, given the strict limits of the legislation regarding data confidentiality under which it operates. IES/NCES legislation protects the privacy of individuals, making wrongful disclosure a Class E felony punishable by up to five years in jail and a $250,000 fine. NCES has experience in working with individually identifiable data through its various sample surveys, and has created the structures and procedures necessary to prevent unauthorized disclosure of such data. In fact, there are no known cases where individually identifiable data collected by IES/NCES have been wrongfully disclosed by

an employee, contractor, or restricted licensee, or of cases in which hackers have breached IES/NCES firewalls. In addition, IES/NCES is the only component of ED that is separately certified by the Chief Information Officer and the Inspector General for its computer operations and system firewalls. Therefore, if collected, the data would be technologically protected and secure.

In the proposed UR system, data would be submitted to NCES but would not ever leave NCES unless authorized by legislation. Estimates created from the UR database would be reported only as aggregates at the level of institutions or groups within institutions. In order to move forward with a UR system, decisions would be necessary regarding which redisclosures are valid uses of the UR data and therefore should be authorized by law.

Burden

The potential for additional burden in terms of labor and financial costs is another challenge to a federal UR system. Postsecondary institutions believe that they are already stretched thin by requirements from federal agencies, state governments, and other organizations to submit data electronically throughout the year, as well as the need to respond to questions in accordance with the Freedom of Information Act (FOIA). At some schools, the need for compliance reporting has used scarce resources that could have been spent on other needs, such as policy analysis. At the same time, many small colleges do not currently have an institutional research office, and any submission of data represents a challenge. Institutions might pass the costs of additional reporting along to students and families in the form of price increases. Nonetheless, when examining the burden a national UR system would present, it is important to focus on the incremental burden of such a system—i.e., the costs produced over and above the costs that would have occurred without the UR system, or to what IPEDS might evolve to if UR is not used. This needs to measure the burden of UR above what might happen to data collection without UR, including the use of XML for transmitting data across computer systems.

The additional burden of a UR system can be divided into two categories: initial implementation and subsequent operations.

Initial Implementation

The burden of initial implementation is expected to be higher than the costs of subsequent operations. The burden may be substantial at some schools that are not already part of UR at the state or private system level. Initial implementation would involve both a field test and the first and second years of full–scale implementation. Generally, there are some challenges that are specific to the initial implementation period, including possible hiring of additional staff or shifting of existing staff, training staff about the requirements of new IPEDS UR reporting, the purchase or upgrade of software, and creating and submitting historical GRS files.

If a UR system were implemented, a field test would be necessary according to IES/NCES standards, in order to make sure that the system works, to anticipate and address problems that would be encountered, and to develop all necessary features in the system prior to implementation. The data submitted for the field test would not have to be complete, but would need to be sufficient to test the system comprehensively. About 1,200 to 1,500 institutions would be required to participate in the field test, and they would have to report using both the old and new IPEDS collection system. NCES would make every effort to notify selected institutions as early as possible. Regardless of the timeframe, it is recognized that participation in the field test would present an additional burden on institutions.

In the full–scale implementation, many institutions would need to upgrade various technologies and assign staff to comply with new reporting requirements. Some institutions would rely on vendors for upgrades to existing software, or to build their UR extracts, which would increase software costs. Others would need to pay for changes to legacy information systems. Staff would need to be trained in the use of these systems and the details of reporting procedures. Institutions in the middle of converting or upgrading their administrative information systems would have particular constraints with implementing these changes during this time. The initial burden on small institutions that may not have sophisticated software or that do not have institutional researchers on campus is likely to be relatively high, unless the institutions are part of a larger system or state association.

Institutional researchers at one TRP meeting argued that obtaining historical GRS files for all cohorts in the first year would present a substantial burden, even though these same files are needed now to calculate the GRS locally. The historical files would not need to include CIP codes or data for time to degree, financial aid, prices, and other information. Required data would include the year of established cohort; first–time, full–time beginning students in the fall; degrees/awards received; as well as a measure of whether students are prepared to transfer. Although institutional researchers attending the TRPs stated that they would prefer to submit historical data the old way if UR were implemented, it was recognized that this would delay the benefits of UR for six years, with no improvement over the current GRS measures. If historical files were submitted, then better calculations such as transfer–out rates at community colleges could begin to be implemented much sooner.

Subsequent Operations

The additional costs of subsequent operations under a UR reporting system are expected to be lower than the costs of initial implementation. Keyholders would need to coordinate with offices on campus to gather data, run internal checks to make sure the data make sense, submit data to NCES several times per year, and work with the IPEDS Help Desk to reconcile record mismatches and discrepancies in the data.

While the data files are being edited, the data would reside in the IPEDS collection system where only the keyholder (and his/her proxies) could view the data. Only after the data were locked, transported to permanent storage, and later migrated to the PAS as aggregate estimates (about a month later) would anyone outside the authorized participants in the submitting institution and the authorized participants in the NCES data receipt process be able to view the data (see the following chapter for details of this process).

Some mismatches of records found by NCES and sent to the keyholder for resolution could be difficult to resolve. For example, NCES would no doubt discover cases in which a student reported by an institution as a first–time student has a match at another institution, indicating that the student had been previously enrolled. NCES might also find a student who matches with other records on all identifiers except last name.

These student records could be sent back to the keyholder to have the mismatch resolved or the identity verified. For these records, the keyholder would then have to check institutional records, resubmit the data if appropriate, or send the data back with no change. These steps might be minimal for each record, and manageable for a small institution. However, the mismatches could add up to a large amount of effort if there are numerous records to resolve, as in the case of large public systems, where the chance of keystroke error might result in many mismatches. Keyholders would have the opportunity to sign off on a mismatch even if they cannot find the student or otherwise resolve the issue. Eventually, the institution's administrative information system would likely be improved to ensure that all students with different types of data can be matched—for example, all students receiving aid would have enrollment records or all graduates would have enrollment records in the system.

Changes in student records during a semester/term, especially regarding changes in attendance status, would require a mechanism for ongoing updates to the UR system. For example, if a student changes from full–time to part–time attendance status, this may affect his or her eligibility for financial aid or loan deferment. This may be difficult for institutions that currently do not store changes in status on their administrative information systems. Similarly, institutions might want to resubmit data from the previous term or year. For example, some institutions post degree awards retroactively during the year after the official degree date, due to incompletes and other issues. This process is available only for the previous year under the existing IPEDS system. The burden of these types of edit failure resolutions and mid–period adjustments could be compounded if institutions are required to keep records of all student data provided to NCES indefinitely in order to remain in compliance with FERPA. If a UR system were authorized, design TRPs would be necessary to decide on the process for dealing with changes in students' records during a term, whether by including mid–term changes in a subsequent file or some other mechanism.

It is difficult to offer cost estimates with any degree of precision, although some rough estimates are presented in Appendix B for purposes of illustration. Costs would likely include additional staff, as well as financial resources for hiring and training, and the costs of software upgrades. These, and other, costs are likely to differ widely among postsecondary institutions, depending on whether they are in state UR systems, whether

they currently upload data to organizations such as the National Student Clearinghouse, whether they use local or proprietary administrative information systems, and whether they currently have relatively low levels of IT and institutional research capability.

At the same time, it is important to keep in mind the labor hours and cost of continuing the current student–related IPEDS components, and the additional costs of expanded data collections in the context of accountability. It is likely that, without a UR system, there would still be increases in the burden of reporting IPEDS aggregate data, including new variables for the construction of net price and perhaps revised graduation rate variables. There would be a corresponding decrease in burden after the initial implementation of a UR system, as postsecondary institutions would no longer need to track and maintain records on GRS cohorts for six years. Rather than each institution calculating its own estimates, NCES would calculate official fall enrollment, graduation rates, financial aid averages, and other measures.

Some institutional researchers attending the TRP meetings noted that the costs to institutions of moving to a UR system would be offset by the benefits of receiving information about students at that institution produced by a UR system. In particular, TRP attendees felt it would be useful to find out, on a UR level, what happens to the students who leave the institution. If such a redisclosure were authorized by the legislation creating a UR system, it would be possible to provide student information back to the institutional keyholder or coordinator. As noted above, issues regarding privacy and the right of students to withhold personal information would need to be resolved in the design phase of the UR system.

TRP panelists noted that if a UR system were implemented, it would be important to try to take into account these various issues during the design phase of implementation so as to minimize institutional burden. For example, design TRPs could develop procedures that minimize the burden of edit failure resolutions and mid–period adjustments while allowing accurate data. Institutions would also have the assistance of the IPEDS Help Desk and various training sessions. Financial aid officials, registrars, IT staff, and institutional researchers would be involved in training and in the project design TRPs.

IPEDS reporting is mandatory for all institutions that participate in the Title IV federal student aid programs. Given the administrative burden of this reporting, it is possible that some funding could be used to mitigate the burden. There are different ways to offset the cost and burden of UR. One funding mechanism, Administrative Cost Allowances (ACAs), is used to help defray the cost of administering federal student aid programs.[34] A similar funding mechanism could be put in place for UR. Another possibility would be for a new grant program to be funded to assist in paying for the costs of implementing a UR system.

Technical Challenges

Technical issues were also raised as a potential challenge to the implementation of a national UR system. The proposed system would include the creation and maintenance of a database of millions of student records, with new records added every year. In addition, the system would require the uploading of large files from postsecondary institutions to NCES, using highly secure mechanisms. Multiple forms of security would need to be in place to protect against unauthorized disclosures of data.

NCES currently has most of the hardware and software necessary to implement a UR system. Much of the equipment used in the current, web–based IPEDS collection would be applicable to a UR collection framework. In addition, ED has servers capable of storing large amounts of student data, as seen for example in the data collected on financial aid applicants by OPE. One necessary addition would be database storage, to be located off–line in a secure site and protected by physical and software firewalls.[35] (See Chapter 4 for more details on software and hardware needs.)

There would likely be greater technical challenges for postsecondary institutions. The extent of the challenge would differ between the registrar, institutional research, and financial aid offices, which sometimes utilize different and incompatible information systems that cannot communicate with each other and complicate the exchange and compilation of data at a central point at the institution. Institutions using both legacy and

[34] Institutions currently receive over $150 million in Administrative Cost Allowances (ACAs), which can be used toward the administration of federal programs such as Pell Grants and campus–based aid.
[35] The cost to NCES would depend on the final design and implementation.

proprietary student information systems would need to make software conversions or updates, while institutions that do not have such systems would need to implement the UR requirements in another manner. NCES has suggested that it can provide an Excel template that could then be used to collect data and generate the data file needed for submission. Although the technical issues could present a problem, these schools currently find a way to do uniform reporting for FSA financial aid eligibility, NSLDS loan deferment, and IRS tax credits.

The proposed UR system envisions the use of XML technology for the submission of data files to NCES, although it is likely that ASCII files would be accepted in the early years of implementation. Some postsecondary institutions have already adopted XML and are using it in their exchange of data with other organizations. For example, in recent years, there has been movement by the Department of Education toward using XML as part of its Common Origination and Disbursement Initiative (COD).[36] FSA has already mandated that institutions begin submitting data to the office using XML by 2005–06. Like FSA, IPEDS is moving to register all of its data elements and collections with the Postsecondary Electronic Standards Council.[37]

Coordination

Coordination of the flow of information presents a multitude of challenges in implementing a UR system. These include the coordination of offices within postsecondary institutions and the management of data collection through the IPEDS coordination trees. The amount of coordination and interaction will be greater due to the increased level of detail that must be matched in the URs.

TRP panelists noted that coordination between various offices on campuses—registrars, institutional researchers, admissions, IT, and financial aid offices—might be difficult, and becomes even more difficult if those offices are running different information systems. Currently, institutions that report data out to states or to

[36] For more information, see http://www.ifap.ed.gov/cod/attachments/CODXMLHandout.pdf. Some TRP panelists reported problems in sending and receiving large files in their work on COD.

[37] The Postsecondary Electronic Standards Council (PESC) is a non–profit association of colleges and universities; professional and commercial organizations; data, software and service providers; and state and federal government agencies, whose mission is to lead the establishment and adoption of data exchange standards in education. For more information, see: http://www.pesc.org/.

organizations such as the National Student Clearinghouse often are reporting a narrower range of data, such as only enrollment data. Merging financial aid data with enrollment and completions data would be especially difficult if an institution does not have an integrated system and uses different student identifiers in its various systems. Further, some schools have decentralized structures with, for example, multiple registrar and financial aid offices. Keyholders would need to coordinate the process of extracting data, editing and cleaning data, and running preliminary aggregate reports across these offices. The problems of editing are minimized somewhat if schools rely on NCES matching subroutines for this first round.

State and system coordinators that are keyholders in the IPEDS coordination tree would have access to collection–level UR data, as they do in the current IPEDS collection system. However, several TRP members argued that a UR system might not work well within the existing IPEDS coordination tree structures. For example, if state coordinators were to continue to submit data to NCES, many would have to expand their collections from institutions or expand the type of data they were collecting. Most state UR systems are based on census dates and analytical files. Some SHEEOs and system offices may not see the benefit of coordinating the submission of multiple streams of enrollment files in the IPEDS UR system unless they get something back, such as the redisclosure of enrollment and attainment data. Yet there would be a noticeable decrease in the requirements for aggregate reporting and a resulting tradeoff in their IPEDS work. This could shift more, or all, of the reporting burden to institutions that had previously relied on SHEEOs to submit data on their behalf. State coordinators might not know how the proposed UR system would affect their role as coordinators until such a system was fully implemented and they could assess the nature of collection demands. The various scenarios for state roles and coordination would be described and documented as soon as possible while the TRP design phase is evolving.

Timing

It is possible that if a UR system were authorized in the Higher Education Act (which could conceivably be as early as the summer of 2005), a field test would then be administered in 2006–07, followed by full–scale implementation in 2007–08. The project

timetable is designed to yield data relatively quickly while avoiding potential problems associated with an expedited timeframe. A phased implementation could also be considered to provide additional time to address problems during implementation.

To respond adequately as part of the field test, it may be necessary for institutions to examine the utility of their administrative information systems for the purposes of producing UR extracts and to address some of the burden issues mentioned above such as training and staffing. Early notification for the selected institutions would be crucial for the institution's ability to respond in a timely and accurate fashion. It is possible that NCES could draw the sample of institutions immediately after legislative authorization to allow selected institutions almost a year to prepare.

If a UR system were mandated, institutions that have not already implemented the two–question format for race/ethnicity, required by OMB Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting on all individually identifiable records, would have to do so during initial implementation.

Another important issue is operational—how to time data collection schedules, while minimizing conflicts and taking into account the treatment of transactional files compared with analytical files. Transactional files reflect operations at a point in time, and usually are used for purposes that require knowing the current status of students, such as enrollment verification. These types of records are not used for the fall enrollment census report that is used to conduct peer comparisons and therefore may not receive as much attention by institutional researchers due to their decreased impact on reporting. Analytical files generally are used for official reporting and analysis, and can be files with a specified snapshot date such as a census date, or files that accumulate all activity during a specified period. The information captured for analytic data reporting may not be the same as a transaction record created for the purpose of student status verification. As mentioned above, many institutional researchers have expressed concern about the burden necessary to clean transactional files before they are submitted. However, it is possible for keyholders to submit files without doing significant data cleaning beforehand and focus attention on the records that come back from NCES as mismatches, which would be resolved by the keyholder before the data were locked and aggregate estimates migrated to the PAS.

To replicate the variables currently collected through IPEDS, the proposed UR system likely would collect enrollment records once per term (with the fall collection file including a census date flag[38] in order to replicate the IPEDS Enrollment component). However, some institutions do not have standard terms; for example, courses may be offered on a rolling basis, or on six–week terms. Flexible term dates are becoming more common with web–based instruction and alternative delivery modes. In addition, some institutions offer courses on both standard terms and nontraditional terms. Institutions might desire to upload data more frequently, especially for the purpose of enrollment verification for student loan programs. The final UR system that is developed through the design TRP process will need to allow for various flexible term reporting options.

Degree and certificate completions would probably be collected with only one file per year. Many institutions have several commencement periods, and might wish to submit multiple files over the year. A problem that occurs with completions data, especially for schools that award degrees only once per year, is that some awards are recorded months after the relevant students have stopped attending institutions. In this case, degree dates reflect the date the degree was awarded rather then when the degree was finished. In designing the timing of data collections and the periods of reference for the data, it would be useful to align the completions data with the enrollment data necessary to calculate graduation rates so that completions records can be matched to comparable enrollment records.

Student financial aid information also likely would be collected with only one file per year. Financial aid data would present some particular challenges in terms of timing, given the logistics of the financial aid award cycle. Data submitted in an academic year would be from the previous year's award cycle, for example. It would be important to time the data collection of financial aid data so that it does not conflict with the institution's aid packaging period, which is the busiest time of year for financial aid offices. In addition, the treatment of summer sessions varies by institution, especially regarding whether summer sessions would trail or lead[39] the submission of an annual data

[38] An alternative to a census flag would be to require a separate census file for enrollment; this option would be considered during the design TRP phase if a UR system were implemented.

[39] Problems will be raised if the file schedule splits the summer session in two; these issues would be addressed during the design TRP phase of implementation. For example, multiple term files can be submitted beyond those that are required.

file. This is an important difference necessary to state resource allocation models that must be considered in a UR system.

All of these timing issues would be addressed during the design phase of UR implementation, should a UR system be authorized. (In addition, a proposed schedule for data collection is included in Chapter 4). In the proposed UR system, collection schedules would not need to be on a uniform schedule, but rather could be geared to a schedule that works best for individual institutions. In other words, institutions with different calendars or financial aid packaging schedules could submit data to NCES on different cycles. Flags would be used to note collection dates and other timing issues. At the same time, the data could all be consistent with the financial aid year, July 1 to June 30. The following chapter delves into more detail regarding the proposed data collection, including the schedule of submission periods.

Variation Across Institutions

All of these challenges are important to the discussion of whether a UR system should be implemented. They are also critical to consider if a UR system is legislated and moves into the design phase. How institutions would be affected by these various challenges depends on their individual circumstances. Institutions that are already uploading student data, that are familiar with procedures, and that possess the required technology would have the fewest challenges in implementing UR. Institutions with good coordination among offices on campus would also be better positioned to implement UR relatively rapidly. On the other hand, many small institutions with few resources would likely have a more difficult time with reporting, especially those without institutional research staff and technological capacity. It will help institutions if they have a good vendor relationship that is meeting their compliance needs; many private, for–profit institutions, for example, build the cost of compliance into their tuition and fees. Many TRP panelists emphasized that if a UR system were to be legislated and implemented, the more the design phase of implementation takes into account these challenges, the less burden it would place on institutions and others.

Chapter 4—System Architecture

The proposed IPEDS UR system is, by choice, general in terms of the design of the system architecture. It was recognized at the outset of the feasibility study process that any such formal design would need to evolve over time, through extensive dialogue with and feedback from constituent communities, and that this would involve numerous future Technical Review Panel (TRP) meetings if implemented. Nonetheless, for the purpose of this feasibility study the nature of the UR proposal needs to be as clear as possible, in terms of assumptions about what such a system would provide. Policymakers, data users, and institutional representatives want to know as much as possible about what the system would require of institutions and of ED to implement and operate, and how it might work conceptually.

This chapter provides a description of the system architecture that is conceptualized for the proposed UR data collection system. Since the UR proposal builds extensively upon that already in place in many states, a basic description of the state UR model is provided first. The underlying assumptions that are inherent in the proposal for a federal version of a UR system are then discussed, followed by a discussion of the processes for collection and editing for a UR system (illustrated through the use of flow charts). Descriptions of other components of a UR system, such as permanent storage, migration to the PAS, authorized redisclosures of UR data, training, Help Desk support, software, and hardware, are also presented.

The State Unit Record Collection Model

The proposed collection of IPEDS student UR data parallels closely that already in place at many SHEEO and state system organizations across the country. These offices collect individually identifiable enrollment, course, financial aid, completions, human resource, room inventory, and finance data for a wide variety of purposes. Any effort to build a federal IPEDS UR system would leverage the important lessons and experiences of the SHEEO and system offices. The processes and end results are similar, though proposed on a larger scale.

Many institutions are required by their coordinating SHEEO or system offices to submit data files with specific variables and census dates. These files are usually uploaded via the web, and in earlier years were transferred across computers using file transfer protocol (FTP) software. While some of this software is web–based, other offices use standard programs written with statistical packages (such as SAS, SPSS, and STATA) to merge, edit, and aggregate the data. SHEEOs and system offices provide software for schools to run edits against their UR data, looking for outliers and missing data. Few of these UR systems look across schools to more accurately classify cohorts and types of students or to share UR data on awards, transfers, and persistence with participating schools.

According to recent information gathered by the national SHEEO association, at least 23 states and/or state systems use their UR data collections to generate and submit aggregate IPEDS reports for member institutions to NCES (L'Orange 2004). Five offices in Alaska, California, Minnesota, North Dakota, and Wisconsin produce all four student–related IPEDS components with their UR systems. Another 13 states/systems reported that they produce at least three of the components this way; while only six states/systems produce the student financial aid IPEDS component from their UR systems.

Some SHEEOs require schools to submit both state UR and aggregate IPEDS reports about enrollment, financial aid, and completions. In these systems, there is a negotiating process that occurs in resolving edits and addressing errors. It is recognized that while the SHEEO/state system editing process and collection data elements might vary from that proposed for IPEDS, NCES would aim to make the two as seamlessly comparable and compatible in definitions, timing, and specifications as possible. If a UR system were implemented, NCES would hold TRP meetings specifically to examine and build upon the effective practices of SHEEOs and state systems in collecting student UR data to produce aggregate IPEDS reports. The relationship between federal and state/system reporting and data structures is likely to develop over time, depending on which mechanisms work most effectively.

Assumptions about System Architecture

The following assumptions about UR system architecture are embedded in the proposal as it has evolved through the feasibility study process. These help to clarify aspects of the emerging design.

- The UR system would work much the same way IPEDS does now. The basic IPEDS collection system would remain the same, involving the secure, online submission of data by officially designated IPEDS keyholders.
- The IPEDS coordination tree would remain in place. IPEDS keyholders would be appointed by institutional CEOs and could appoint proxies to assist in submitting different types of data. For schools that fall under an IPEDS coordination tree, two or more levels of review and approval would still be in place—once a school locks its data, the coordinator would review them and begin either to edit or lock the file. The process would work the same for URs as it does for the aggregate IPEDS components.
- Data edits would be customized to each school. As is the case now, schools would only submit data for items that pertain to their mission and type, as defined through the Institutional Characteristics (IC) component.
- Editing would be done as before, but expanded. Currently, for example, the previous year's aggregate data are compared with a new submission to look for consistency; if data are different than the expected range, the keyholder is asked to review the results. This type of edit review would continue for URs, but with more levels and types of edits to be resolved and passed.
- Once filed, the UR data would be physically transported to permanent storage. After the UR and aggregate level data passed edits and moved through the coordination tree, they would be physically moved from the collection system to a special UR database.

- Estimates from the UR database would be used to populate the Peer Analysis System (PAS). Currently, when aggregate IPEDS data are moved from the collection system, they are stored in the PAS. The PAS data are then used to load information into the IPEDS COOL website. Although the UR system would not change this general process, data would come from the new UR database rather than the collection system. Aggregate estimates would be calculated, edited, locked, and if necessary perturbed before they were moved to the PAS, per IES/NCES statistical standards and requirements to protect confidentiality. (GRS and SFA data would be perturbed; enrollment and completions data are not subject to perturbation).

- Permissions and levels of access in the PAS would remain the same. In the existing IPEDS collection system, data are migrated to the PAS and are made available at what is termed the "collection level" several weeks after submission, so that keyholders may immediately begin to compare their institution to others using unofficial, preliminary data. After extensive cleanup by NCES, the data are made available to other users at what is termed the "institution level." Only after the data are finalized and made official through adjudication are the data moved to "guest level," which is available to the general public. These three levels of access are available in the PAS. Although the manner of collection would be different if UR were implemented, the release of aggregate IPEDS data through the PAS would continue under this system of collection, institution, and guest level access.

Collection System

Generally, the UR collection system would be designed to collect individually identifiable, student–related data through files that are submitted electronically by institutions. The files would be used to calculate institutional summary totals for each school, with information about enrollment, completions, graduation rates, financial aid, and price. A list of the data elements that would probably be needed is presented in table

1. These elements are listed by file type, along with the IPEDS component or other federal mandates that require them, and information about their format.

Institutions would be expected to submit the four types of files: 1) header files, which would provide individually identifiable information such as name, Social Security Number (SSN), date of birth, address, race/ethnicity, and gender that are attached to an individual student's record; 2) enrollment/term files, which would include program information such as number of courses and credits attempted, major field of study, start and end dates, and attendance status; 3) completions files, which would include information on degree completions and the date of completion; and 4) financial aid files, which would include information on financial aid received from federal, state, and institutional sources, as well as price of attendance. Each uploaded file would include a single record per student, per term/reporting period, per institution. Each of the term, completions, and financial aid files would need to include the same header information which is needed to match records across files; including fields such as SSN, name, date of birth, gender, and address.

In addition, in the first year of collection, historical enrollment files on Graduation Rates (GRS) cohorts would be required in order to complete the required calculations for the GRS. Institutions would be expected to submit the four types of files over the course of a year, much as they currently submit the aggregate IPEDS components, depending upon when the data become available. Some files would be submitted once a year, while others would be submitted more frequently (see below).

Instead of filling out online screens, keyholders would upload files in text or AASCII format, eventually in XML. Keyholders would view their aggregate reports for the EF, SFA, GRS, C, and IC price components online and review edits. In place of submitting corrected reports, however, they would upload corrected UR data, viewing the report results until they pass edits and are locked.

Schedule of Data Collection

The dates for file submissions would be set in order to keep enrollment, completions, financial aid, and graduation rates in line with the student financial aid year. This would ensure that reporting across these components is comparable. This would also

establish a consistent IPEDS year that would be in conformity with the ED FSA year for financial aid, which is July 1 to June 30. As a result, IPEDS—which is mandatory for Title IV institutions—would operate on the same cycle as FSA.

Table 1. List of proposed variables to be collected in an IPEDS unit record system, by file type and IPEDS requirement

	File type	Requirement	Other comments
Student name	Header	Matching	Full name, not just first and last names
ITIN or SSN	Header	Matching	
Permanent address	Header	EF (Residence and Migration)	
Date of birth	Header	EF (Age)	
State of residence	Header	EF (Residence and Migration)	
Gender	Header	EF, C, GRS	
Race/ethnicity	Header	EF, C, GRS	OMB two-question format
Citizenship	Header	EF, C	
Program	Header	EF, C	CIP 6-digit, including 'undeclared'
Degree plan	Header	GRS	
Program length	Header	GRS	Needed for 'long' programs
Varsity sport	Header	GRS	
High school graduation date	Header	EF	Needed for coenrollment
Institution UNITID	Header	Internal upload	
Transaction date	Header	Internal upload	
Redisclosure flag	Header		
Start date	Term		
End date	Term		
Number of courses	Term	EF, GRS	
Credit hours	Term	EF, GRS	
Attendance intensity	Term	EF	
Level	Term	EF	Including first-time, undergraduate, graduate, first-professional
Census date	Term	EF	Or flag for EF inclusion
Transaction date	Term	Internal upload	
Tuition and fees	Aid	Prices	
Total price of attendance	Aid	Prices	
State residency status	Aid	Prices	In-district, in-state, out-of-state
Campus residency status	Aid	Prices	On campus, off campus, with or without family
Dependency status	Aid	SFA	
Federal grants	Aid	SFA	Including grants from ED as well as other federal agencies
State grants	Aid	SFA	
Institutional grants	Aid	SFA	
Loans	Aid	SFA	Including federal loans and loans from other sources
Assistantships	Aid	SFA	
Transaction date	Aid	Internal upload	
Degree granted	Completions	C, GRS	
Degree date	Completions	C, GRS	
Exclusion flag	Completions	GRS	
Ready for transfer	Completions	GRS	
Transaction date	Completions	Internal upload	

NOTE: Term files contain enrollment information. IPEDS components are the following: ITIN = Individual Taxpayer Identification Number; SSN = Social Security Number; EF = Enrollment; C = Completions; GRS = Graduation Rates; OMB = Office of Management and Budget; CIP = Classification of Instructional Programs; UNITID = Institution Identification Number; SFA = Student Financial Aid; Prices = Price of attendance.

IPEDS would use the July 1 to June 30 year for full–year enrollment, credits attempted, completions, student financial aid, and graduation rates. The enrollment transaction files could be selected as appropriate to accommodate these dates. The student financial aid data would be identical in dates to those used for Fiscal Operations Report and Application to Participate (FISAP) reporting. It is recognized that the completions and

graduation rate data might result in slightly different calculations if aggregates were to be computed this way, but only for the first reporting period.

If this schedule were put in place, it would require modification of the census dates currently in use for the Graduation Rates component to match those used for financial aid. These dates would be used as the computational year and would not necessarily be the same as the academic year used by institutions. The nominal time for calculating the attainment of a bachelor's degree is based upon a fall enrollment and a spring graduation, not on a late summer graduation. The calculation of a graduation rate that is 100 percent of nominal time would be based on completion within three years and nine months, not four full years, since most bachelor's degrees are conferred in May. This change in GRS calculation dates would require a revision to the Student Right–to–Know legislation as part of any Higher Education Act reauthorization. The first year that the GRS was calculated with URs, the rates would be slightly different because of this shorter time period. This would return to a standard period of time the following year. These rates would then be comparable to those calculated for time–to–degree using the completions and header data. The change in graduation rate timing would affect trend data, and would likely create problems with reporting study abroad and summer completers in the first year.

Header records would be submitted at least once for every student, so that all students have at least one header record in the UR system. These would be replaced only when there is a change in the student information. Another option that emerged in the feasibility study is for schools to submit these records each term, over and over, replacing those previously submitted. Regardless of which option was implemented, it would have to ensure that certain key information changes were documented. These files would provide the information needed for matching the different types of files and would therefore include social security or taxpayer identification number, first and last name, gender, date of birth, and permanent address.

The enrollment/term files would capture intensity of instructional effort for purposes of financial aid and loan deferment eligibility. Student level, courses attempted, and credit hours attempted would also be documented. Since the enrollment files would be used to generate the fall enrollment data, the records for those students that should be

included in the fall census snapshot would need to be flagged.[40] The schedule for submitting term enrollment files would vary depending upon the institutional calendar as documented in the Institutional Characteristics component. Typically, the data would be submitted quarterly for schools on the quarter system and three times a year for those on the semester system. All files would be merged into a composite file that documents all enrollments for the 12 months between July 1 and June 30. For schools with a quarter system, for example, files would be submitted for the enrollment periods July–September, October–December, January–March, and April–June (see Table 2 below). In some cases, the summer term would lead the files for a given year, while in others it would trail at the end of the year, and both could possibly be permissible as long as this was documented in the IC file. The treatment of the summer term might vary depending upon the institution's financial aid year for the submission of files. Institutions with continuous enrollment would need to submit files at least quarterly.

Table 2. Schedule of data collection

	Enrollment			Completions			Financial aid		
	Submission of files	Data lock	Migration to PAS	Submission of files	Data lock	Migration to PAS	Submission of files	Data lock	Migration to PAS
July	Submit April–June file								
August									
September		X		Submit file (July to June of previous year)					
October	Submit July–September file (including census flags)		X				Submit file (July to June of previous year)		
November					X				
December		X				X		X	
January	Submit October–December file (including census flags)		X						X
February									
March		X							
April	Submit January–March file		X						
May									
June		X							
July			X						

NOTE: PAS = Peer Analysis System.

Regardless of how often they were submitted and for what dates, the files would need to capture all activity within their official time period, and include flags that identify

[40] An alternative to a census flag would be to require a separate census file for enrollment; this option would be considered during the design TRP phase if a UR system were implemented.

which students should be counted in the fall census and enrollment report.[41] The files for July–September and October–December would include all students who should be counted in the official fall enrollment census date of October 15 that is used for the IPEDS EF component.[42] These students' records would be flagged to document that they should be included in this aggregate report. The October–December file would have to be filed and locked by the end of March at the latest, so that it could be migrated from the collection system to the PAS by the end of April. This timeframe would be possible because of the extensive editing and matching processes that would be built into the submission process.

The proposed schedule of file submissions would provide more flexibility in term reporting than is the case under the existing IPEDS components, revolving around a schedule that works for each institution. With this system, schools would have extended time to ensure that their aggregate reports pass edits and that their aggregate EF reports are accurate (see details about the edit process below). Data would be provided in the PAS at the end of April, whereas enrollment figures are currently being released at the institution–level at the end of May. The use of UR would help speed up the production of EF data, which are needed by schools to document official IPEDS enrollment figures for internal and external purposes.

Completions file(s) would contain the date, type, and level of degree or certification award. In cases where transfer was the measurable outcome, the file would include whether the student is ready for transfer. The file would also include any exclusion codes that are required for maintaining the correct GRS cohort. Institutions would be given two choices for submitting completions data. With the first, they would submit one annual file at the end of September for all awards between July 1 and June 30. As an alternative, they could upload individual files for each commencement during this time period. All files would still be due in September. Once the data are locked in November, they would be used to populate aggregate reports. Cleanup would be done and the data would be migrated to the PAS at the collection–level in December.

---

[41] It is important, whatever number of files are submitted or schedule is adopted, that the composite result include all possible instructional activity within a year's time, so that aid eligibility and loan deferment status could be determined in a timely manner and graduation, transfer, and persistence measures could be calculated accurately and with more precision than is currently the case.

[42] The two sets of files would include the records of all students who started a term that would be included in the October 15 census data count.

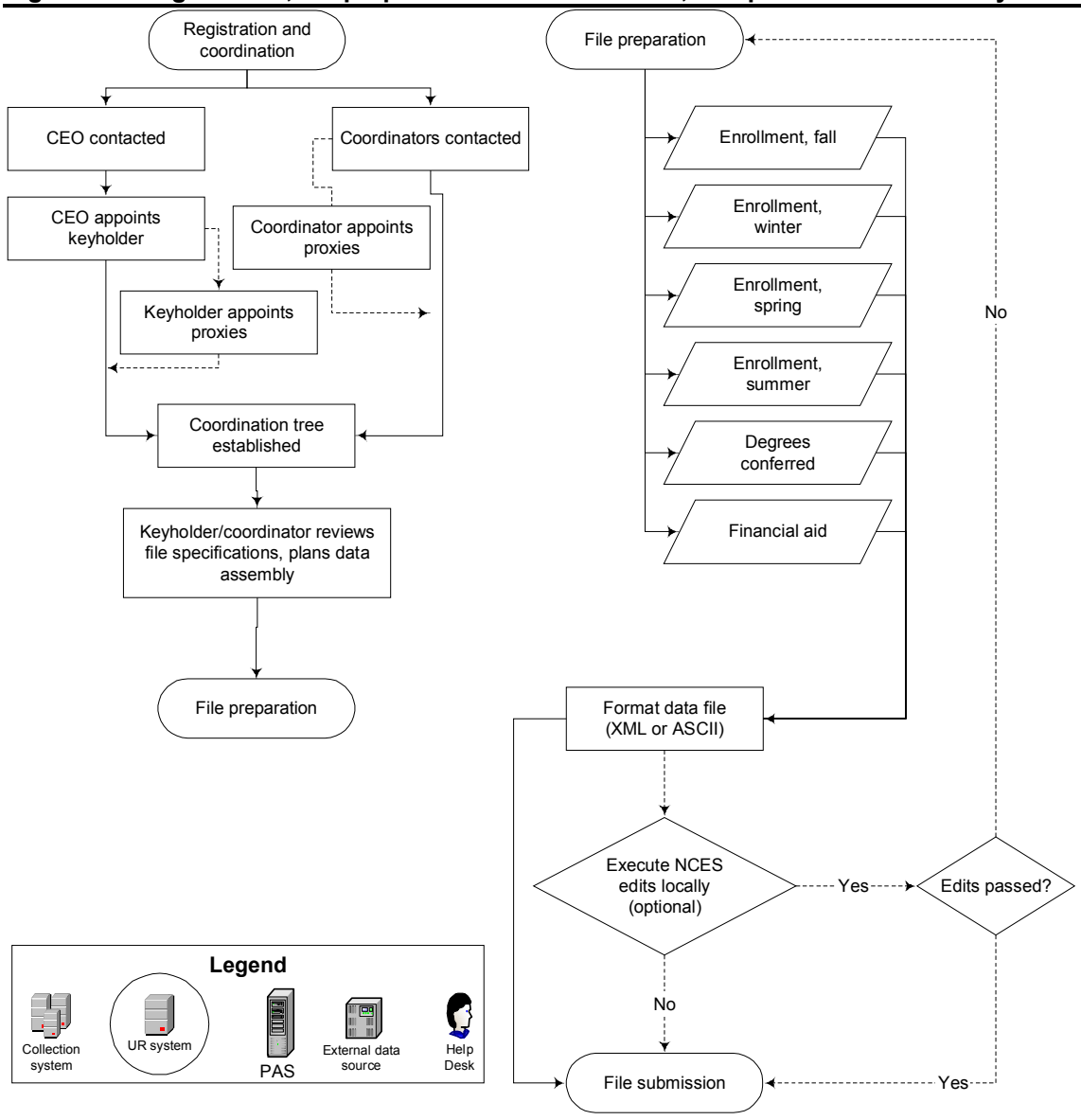
Price and financial aid data would be included in the aid file, along with dependency status and other fields necessary to accurately calculate the net price paid by the individual student. Given a financial aid year from July 1 to June 30, institutions would submit one annual file that includes all financial aid awards during this time period. The file would be uploaded at the same time that FISAP is due to FSA, on October 1. The file would be locked in December, processed with internal NCES edits, and net price information would be posted in January. Tables for IPEDS COOL would be created and posted in February. Although these data would be one year old, the financial aid data are needed in great detail for the calculation of net price. The current year's published price would be collected as part of the Institutional Characteristics component in order to maintain more recent figures.

In addition to the annual schedule of file submissions, there would need to be several data submissions as part of the initial year of implementation of a UR system. First, institutions would need to supply header records for all students, not just new students. Second, historical information on GRS student cohorts would be needed in order to perform the multi–year calculations needed for graduation and transfer rates. For example, four–year institutions would be required to upload up to six years of historical GRS data. Data would be needed for all established GRS cohorts for the different program lengths at each institution. The full, historical student enrollment and financial aid records for each cohort would not be necessary; only the current header record for every student in any active cohort, including those still enrolled, completers, and known transfers. The GRS cohort header record would include several fields not in other header records, such as first date of attendance, current status (enrolled, completer, or known transfer), date of most recent status change, and any exclusion flags that may apply.

File Preparation and Submission

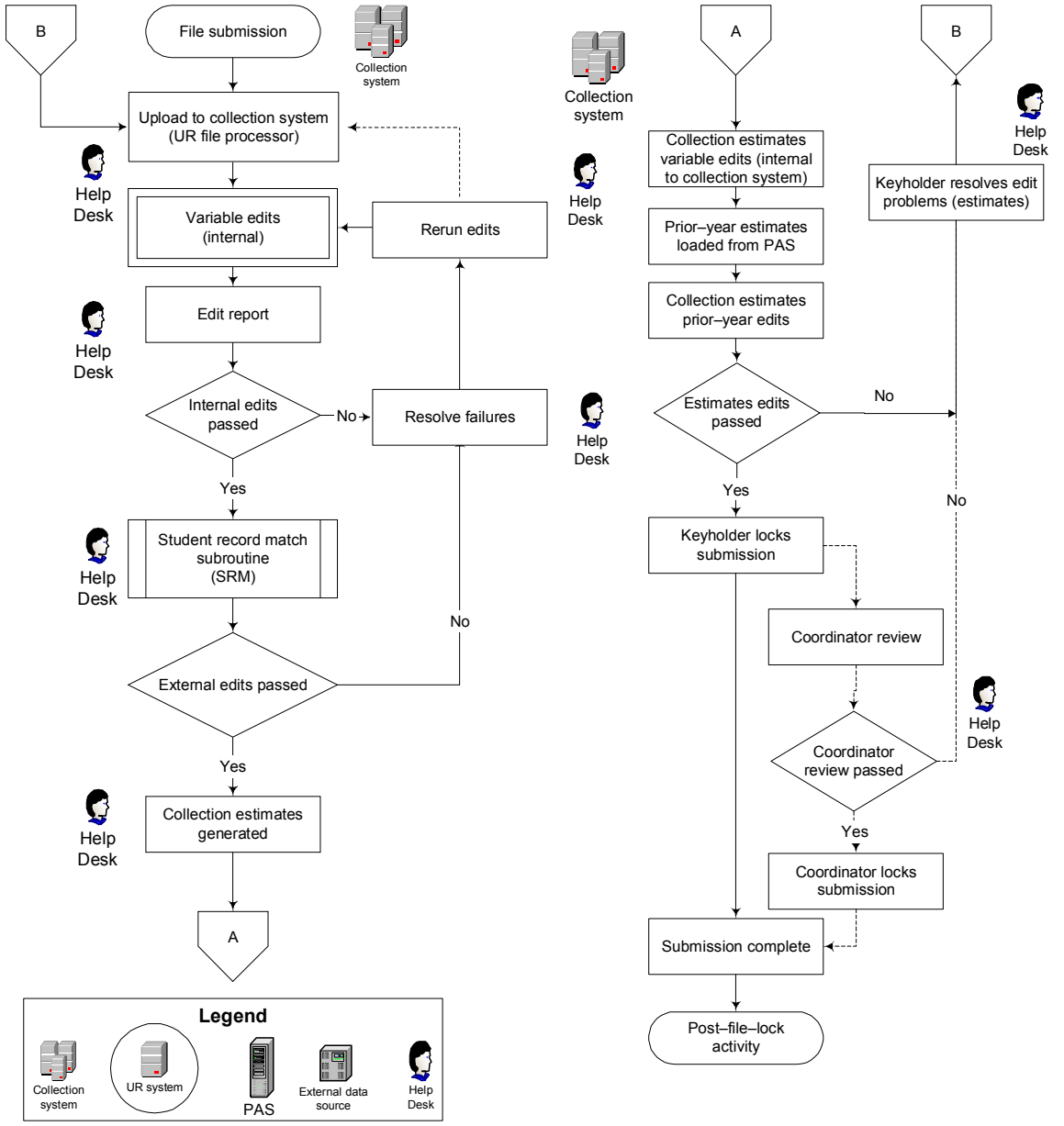
A flow chart of the collection system follows that depicts the registration and coordination processes for the IPEDS coordination tree; file preparation before submission; online file submission; resolution of edit problems; locking of data files by the keyholder; and subsequent activities within the IPEDS collection schedule (figure 2). These stages are explained below in more detail.

Figure 2. Registration, file preparation and submission, and post–file–lock activity



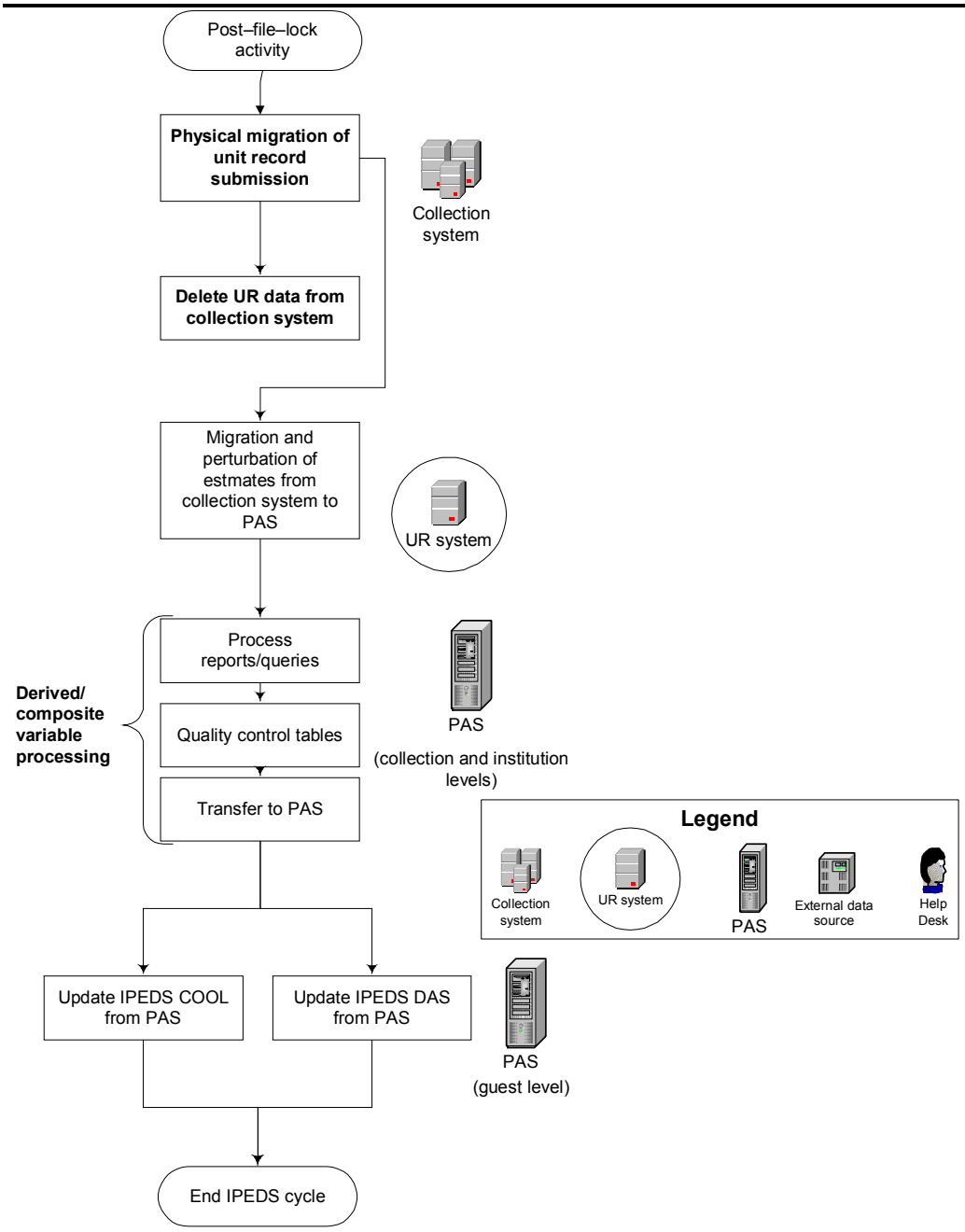
See notes at end of figure.

Figure 2. Registration, file preparation and submission, and post–file–lock activity — Continued



See notes at end of figure.

Figure 2. Registration, file preparation and submission, and post–file–lock activity — Continued



NOTE: Flowchart is for a proposed UR system; if such a system was authorized and funded, the details of the system architecture would be decided after several Technical Review Panels as well as a field test. The collection system icon refers to the IPEDS collection system for aggregate and unit record data collection. The UR system icon refers to the IPEDS offline database unit record data storage and analysis. The PAS icon refers to the IPEDS Peer Analysis System storage and analysis. The external data source icon refers to nonspecific offline data systems. The Help Desk icon refers to the IPEDS Help Desk. IPEDS COOL refers to College Opportunities Online. IPEDS DAS refers to the IPEDS Data Analysis System.

The registration process under a UR system would be similar to the current IPEDS framework. Registration is begun with a letter from NCES to the school CEO, who appoints an official IPEDS keyholder (see figure 2, registration and coordination). The keyholder may then appoint proxies to work with or view different data pieces. For those institutions that fall under an IPEDS coordination tree, the state or system coordinators also are contacted and these persons may in turn designate proxies within their organizations. The appointment of institutional keyholders and their proxies and coordinators and their proxies constitute the establishment of the "coordination tree." All future IPEDS collections must follow this tree for submission and approval or locking of data.

Under a UR system, in preparation for a new collection year, all keyholders and coordinators would review the file specifications and plans for assembling all of the IPEDS components. Then the process of preparing files would begin, including the student–related files for the UR system (figure 2, file preparation). These files would include enrollment, degrees conferred or completions, and financial aid, as well as a header file for every student who has one or more of these three types of records. File preparation would begin with the creation of a data file that conforms to the requirements and specifications from NCES in terms of what data need to be included and how they should be formatted. The data would need to be submitted in XML format, though AASCII text format might be permissible for some institutions during the implementation phase.

The timing and sequence of data files for enrollment, completions, and financial aid would be designed so that institutional researchers and other school staff would have as much time and opportunity as possible to clean up the data before they become public. Obviously, if schools submit completions files more frequently than once a year, there would be additional time spent with these than if one composite file for the entire year were produced.

Excel templates would be provided for schools requiring help in exporting XML data. NCES also would provide the source code that documents in Structured Query Language (SQL) the many edits that would be conducted at NCES once the data are submitted. As a preliminary step, some institutions might want to execute these edits locally on their own computer systems, although it would not be required under the UR

system. If the local program passes edits, then the data would be ready for submission; if not, additional file preparation and editing would be needed. Because institutions would not have access to the complete database, the final IPEDS numbers would differ slightly from those calculated locally. Some institutions would likely choose not to run edits locally, preferring to submit the data files and then deal with the mismatches sent back to the keyholder.

Edit Process

In the next step of the collection process, keyholders would upload a data file to the collection system (figure 2, file submission). A wide range of levels and types of edits would be put in place for URs, and these would all need to be resolved and passed. Any corrections to the aggregate data would be made by submitting new UR files. The focus would remain on producing accurate results for the aggregate EF, C, GRS, SFA, and IC price components. However, the difference between how IPEDS works now and how it would work using URs is that the only way to correct reports under the UR system would be through submitting new URs, rather than submitting another summary file. Also, the editing process would admittedly be more complex, working at both the UR and the aggregate level.

In the first step, edits for internal consistency would be run on the NCES server, resulting in an edit report. The IPEDS Help Desk would work with the keyholder to resolve any failures at this stage, until the data pass edits.

Once the internal edits are passed, the data would proceed into the record–matching phase. In this phase, the data would be physically transported to the permanent UR storage database and matched with other students' records for discrepancies. For example, a school could classify a student as first–time, full–time freshmen, but NCES verifications across the entire student population might find that she/he took previous course work elsewhere. Discrepancies would need to be resolved and re–run until they are passed successfully. The process of matching student records is described below in more detail. After records are matched, then the UR collection data would be ready to be used to generate institutional, aggregate estimates.

In the final stage of edits, collection estimates would be generated and compared with the previous year's aggregate estimates from the PAS (figure 2, collection estimates variable edits). If the comparison of aggregate reports passed the edit process,[43] then the keyholder would finalize or lock the submission and the submission would be considered complete. If a coordinator were involved as part of the coordination tree, then another level of review and approval would be necessary before the submission was considered complete. If the keyholder edits did not pass, or the coordinator review found problems, then the edit problems would again need to be resolved by the keyholder, requiring the resubmission of the data.

Movement of Data to Permanent Storage and Aggregates to PAS

After the UR data pass edits, move through the coordination tree, and are locked, internal NCES processes would occur before the end of the collection schedule (figure 2, post file lock activity). The data would be physically transferred from the collection system to a special UR database. This would be done separately for each institution, so that UR data would reside on the collection server during only the brief time in which submissions and edits were being made. All UR data would then be deleted from the collection system. The permanent storage database would reside behind all firewalls and would not be connected to the Internet; it would not be connected directly to the collection server or to the database server that would house the PAS. Data would be transferred manually through media by secure NCES staff. No one outside of these approved NCES staff and contractors would have access to the permanent storage database for UR.

As part of the move from the collection system to the UR system, aggregate estimates would also be moved to the PAS at the collection level. Aggregate data from SFA and GRS would be subject to disclosure risk avoidance analyses and would be perturbed where required by IES/NCES policy to maintain confidentiality. At this stage, a variety of quality control reports and analyses would be conducted using the migrated PAS data to ensure their data integrity. The collection level estimates would be available

[43] In the PAS, keyholders would view aggregate data that have gone through the NCES edit process; redisclosures of NCES edit changes (including perturbation) would not be made to institutions.

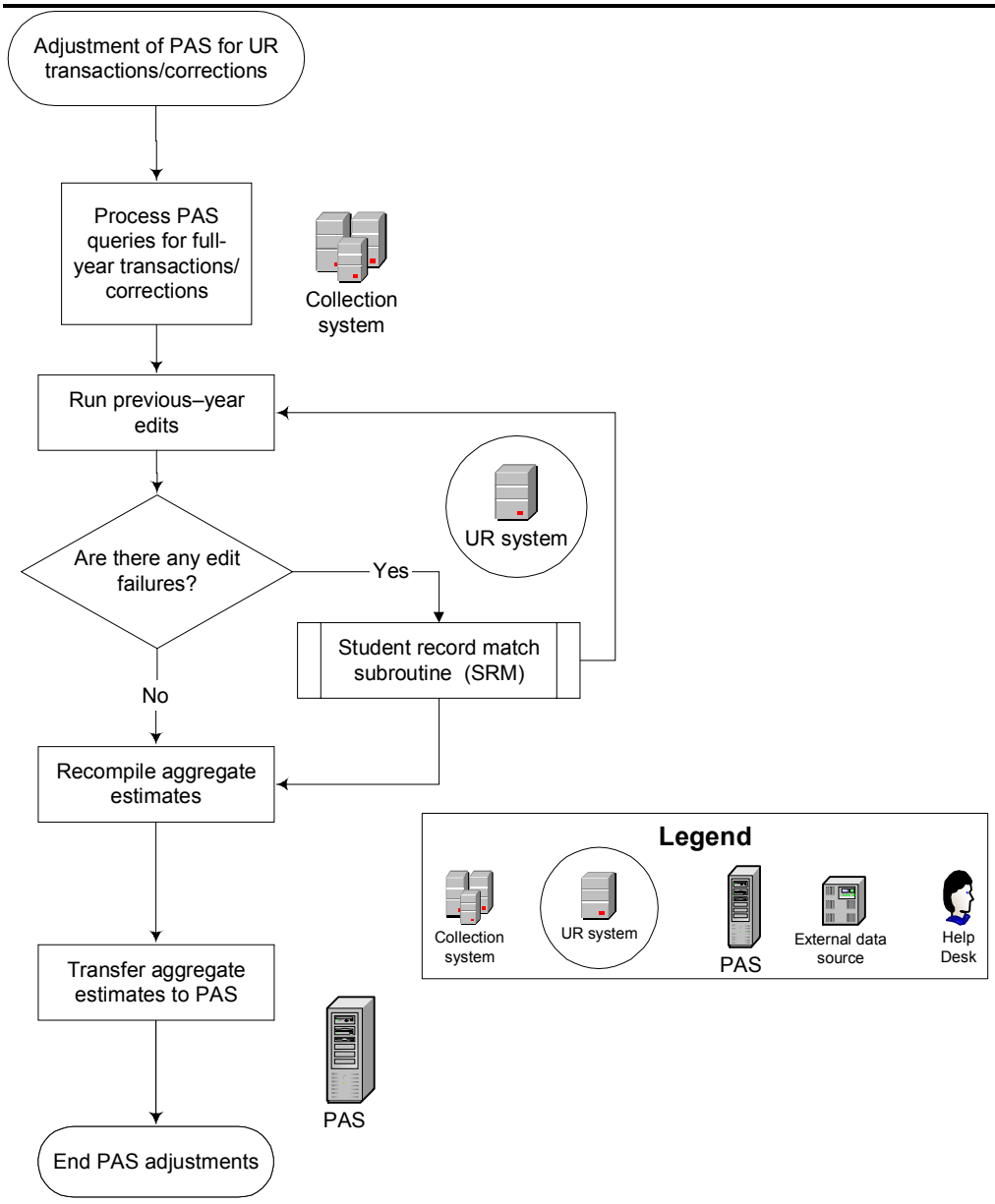
only to keyholders and their proxies at this point. After passing these quality controls, the estimates would be moved to the institutional level of the PAS, where they would be available to institutional level users for peer comparisons.

As multiple UR files are submitted, new data would be made available after they go through the editing process (subject to the advice of design TRPs). Different estimates of aggregate totals might be calculated for institutional data, requiring adjustments to the PAS for UR transactions and corrections during the year (figure 3). These could be combined into a full year's worth of transactions with various corrections for header information and other changes. Previous year edits would be run. If edits failed due to corrections, then the student records would be put through another student record matching process. If they passed edits, then aggregate institutional estimates would be recompiled and then transferred to the PAS. As a general rule, no more changes would be permitted until the following year, when schools would be able to resubmit prior year data to make corrections. This is one proposed model for dealing with mid–term adjustments; design TRPs would need to explore this issue to determine which mechanism might minimize burden on institutions while allowing necessary adjustments such as attendance intensity for enrollment verification.

Once the data were finalized through adjudication, the estimates would be moved to the guest level of the PAS and would be used to update two additional data tools, IPEDS COOL and the IPEDS DAS. Adjudicated, final files would follow per the required, standard IES/NCES review and approval process. With the availability of data on the IPEDS PAS, COOL, and DAS, the IPEDS cycle would end. Aggregate reports in the PAS could be expanded if additional derived variables were to be raised in design TRPs and approved by NCES. In this case, the new variables would be available at the collection level of the PAS.

Institutions would be able to view aggregate reports for each file at each stage of the collection process, including after the files were locked. Generally, it would take approximately one to two months to do cleanup of the data after they have been locked. It would take approximately two months after the end of the collection period for data to be moved to the institution–level access in the PAS. This timeframe is comparable to the current IPEDS workload for institutions and would need to be reviewed and refined as

Figure 3. Adjustment of Peer Analysis System for unit record transactions and corrections



NOTE: Flowchart is for a proposed UR system; if such a system was authorized and funded, the details of the system architecture would be decided after several Technical Review Panels as well as a field test. The collection system icon refers to the IPEDS collection system for aggregate and unit record data collection. The UR system icon refers to the IPEDS off–line database unit record data storage and analysis. The PAS icon refers to the IPEDS Peer Analysis System storage and analysis. The external data source icon refers to nonspecific offline data systems. The Help Desk icon refers to the IPEDS Help Desk.

part of the TRP design phase if a UR system were implemented. Adjudication would continue to take longer to produce guest–level PAS files than many would prefer; however, this process is continually being improved to be as quick as possible given the levels of approval required.

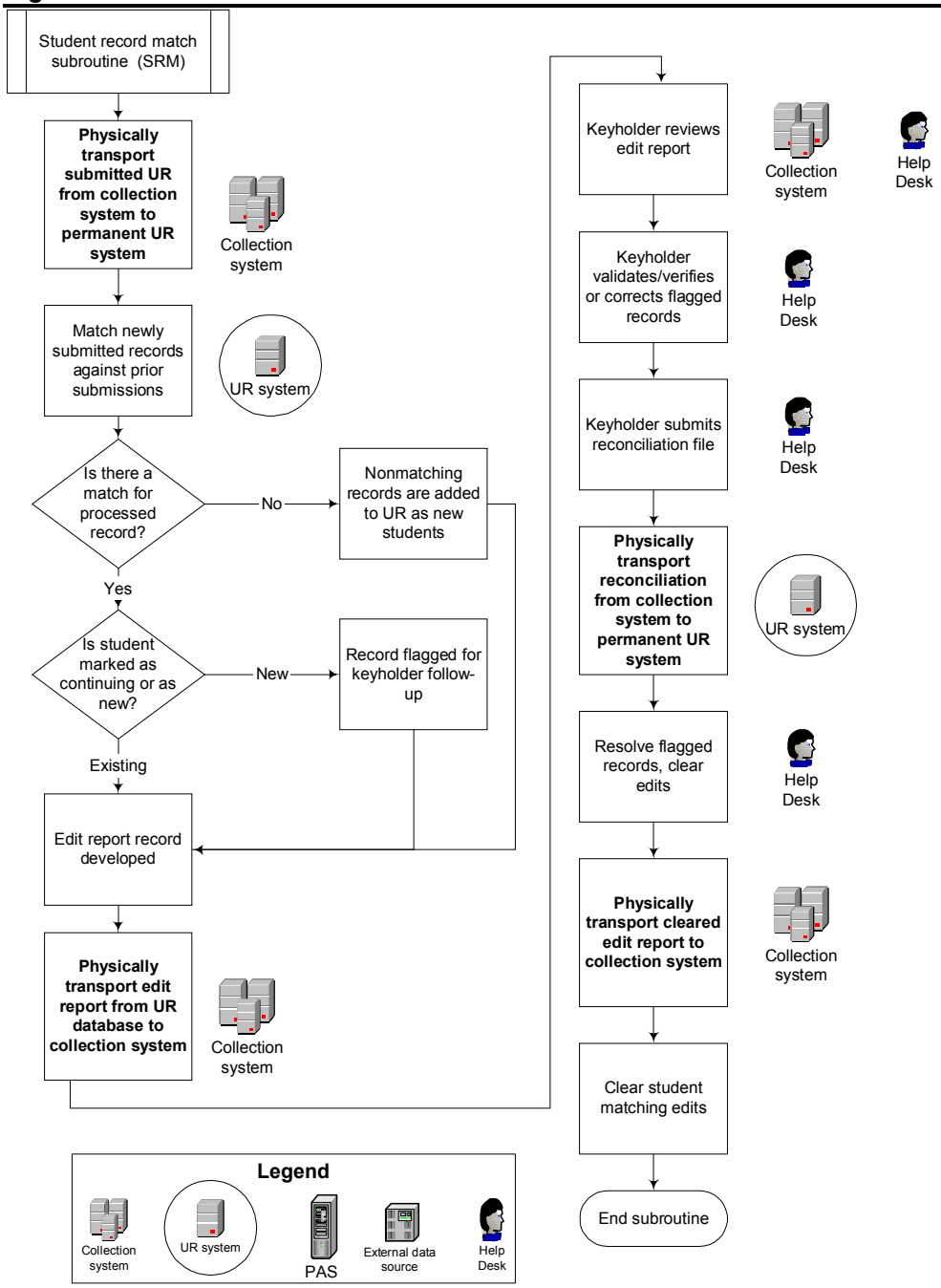
Process of Matching Records

The process of matching student records with the larger UR database might happen at various stages of the collection system. First, matching of records would occur during the edit process after submission of data files by keyholders. Mismatches would be resolved by the school's keyholder working with the IPEDS Help Desk. Special algorithms and "fuzzy matches" would be used to suggest logical, possible matches and how best to resolve discrepancies between records, so that time spent on reconciling mismatches would be minimized. Figure 4 illustrates that in matching student records, data would be physically moved from the collection system to the secure, UR system where they would be matched against prior records and submissions.

When new files entered the system, they would be checked against existing records in the UR database; if there were no match, a student would be confirmed as new to the system. If there was a match, a student record would be confirmed as continuing, unless the previous work was concurrent high school enrollment and the student qualified as a first–time freshmen. The record would be flagged and an edit report would be created and sent to the keyholder for review in the collection system. With the assistance of the Help Desk, the keyholder would either correct a flagged record or verify that a student was new. Keyholders would go through various processes, including reviewing edit reports; validating, verifying, or correcting records; and submitting the reconciled data file. After this, the data would once again be physically transported to the UR system, where flagged edits would be resolved and edits cleared. Then, the data would be moved back physically to the collection system, where all student matching edits would be cleared and the process of matching would be finished as part of this subroutine.

Other mismatches between record sets would need to be verified by keyholders working with the Help Desk, such as a financial aid record for which there was no enrollment record (this could occur for several reasons, including the late post–processing

Figure 4. Student record match subroutine



NOTE: Flowchart is for a proposed UR system; if such a system was authorized and funded, the details of the system architecture would be decided after several Technical Review Panels as well as a field test. The collection system icon refers to the IPEDS collection system for aggregate and unit record data collection. The UR system icon refers to the IPEDS offline database unit record data storage and analysis. The PAS icon refers to the IPEDS Peer Analysis System storage and analysis. The external data source icon refers to nonspecific offline data systems. The Help Desk icon refers to the IPEDS Help Desk.

of an award amendment, which is done after a student has already graduated). A similar process might also occur if the PAS needed to be adjusted for UR corrections.

As described previously, NCES has extensive experience in matching URs from different sources as part of NPSAS. It is estimated that approximately 4 to 6 percent of such file merges result in mismatches in records that must be resolved. For an institution with a headcount enrollment of 10,000 students, there could conceivably be 10,000 header records and 10,000 term enrollment records for a given file submission. At a 4 to 6 percent mismatch rate, this would result in approximately 400 to 600 records that would need to be resolved. However, there would be plenty of time for schools to edit or validate the matches for these records within the schedule of file submissions.

The actual increase in burden with the implementation of URs would be the time necessary for resolving these 400 to 600 records. Some institutions consider estimates of the time staff would spend merging records and creating draft summary reports locally to be part of the additional burden of a UR system. Many TRP members believed that the student matching routine would be time consuming and difficult. However, this burden of verifying the outcome and matching records locally would not be designed or required as part of the UR system, but rather would be a choice of institutions to conduct as an additional process. During the TRP design phase, NCES would work to ensure that there is adequate time for editing and matching for submissions.

Issues in the Collection Process

In utilizing the experience and knowledge of SHEEOs and state systems, of TRP meeting panelists, and of others who have provided comments and feedback as part of the feasibility study process, there would be some important data and collection issues to address if IPEDS student URs were implemented. Some of these are highlighted below.

- Working with term dates. The calculation of credit hours/contact hours and instructional intensity would be difficult if it must take into account different types of terms with different start and stop dates. A methodology would need to be developed to allow for some kind of agreed–upon weighting scheme, so that measures of output are comparable and could be converted across types and time frames.

- Extracting data. Some schools might utilize vendor products to extract the data for IPEDS URs about students, financial aid, and completions. Other schools might involve different offices and require coordination between admissions, financial aid, registrar, bursar, and other functions. Each system might have automatic extracting capability, but be designed for different purposes that are predominantly operational in nature versus geared for federal reporting and analysis. The IPEDS UR extracts would need to have identical census dates and business rules for what records to include, so that they could be merged appropriately to produce the required data.

- Problems in merging IPEDS data files. With this system of data files, NCES would need to match records between files for reporting analyses. For example, completions records would be matched against header files to obtain demographic information. Financial aid and completion records would be matched against enrollment files to obtain enrollment eligibility and cohort information. If records in one or more merged files were missing, the error rate would increase and there would be increased need for institutions to work with the IPEDS Help Desk to resolve the mismatches. Therefore, at the outset, training would be provided to ensure that extracts are designed with the merging of data files in mind. Institutions would be given the SQL code with which to do these types of merges. For those that rely on vendors to automatically generate these files in required IPEDS format, there would be software development involved.

- Timing of reporting. Institutions responding to IPEDS would have their official fall enrollment, graduation rates, persistence and retention, and financial aid averages calculated for them by the NCES software. It is important that these data be collected and disseminated in a timely manner so that this process does not impact important deadlines for other types of submissions, such as to SHEEOs or admissions guide publishers. The UR system would be designed so that it ensures that reporting is reasonable, accurate, timely, and meets other reporting needs besides those of NCES.

- Change in the locus of reporting. NCES would provide institutions with the SQL code and data structures necessary to calculate aggregate IPEDS reports locally. Some schools would want to do this before submitting the data to IPEDS. However, since some of the data would be revised based on redisclosures (such as for the corrected determination of first–time, full–time freshmen status), some schools might not be able to replicate this calculation entirely locally, because they cannot access UR data for all schools. This is not very different from the changes that are made as part of data migration from collection to the PAS to preserve confidentiality through perturbation. Still, this represents a potential change in the locus of control for reporting.

- The need for different streams of data. In designing a system that brings together data from different offices on campus at different times for different purposes, one possibility would be to stream these data directly from those offices to NCES. In this case, enrollment and completions data would be submitted by the registrar, financial aid data from the financial aid office, and price data from the bursar. Schools would not have to merge record sets locally, but would work with NCES and the IPEDS Help Desk after submission to resolve mismatches. Institutions would benefit from more sophisticated and peer–reviewed federal edit and matching processes, in order to provide much more accurate information for consumers and policy makers than they would be capable of doing themselves due to staffing and resource constraints. In this model, the streaming of data would help build and strengthen colleges' and universities' capacity for institutional research.

Redisclosures and Other Data Uses

As mentioned in previous chapters, the proposed UR system includes several redisclosures that would need to be authorized by the legislation creating UR. In addition, there are other uses of a UR system that would involve matching or sampling of student

records. In order to address privacy issues, it is proposed that students would have the right to "opt out" of some of these uses.

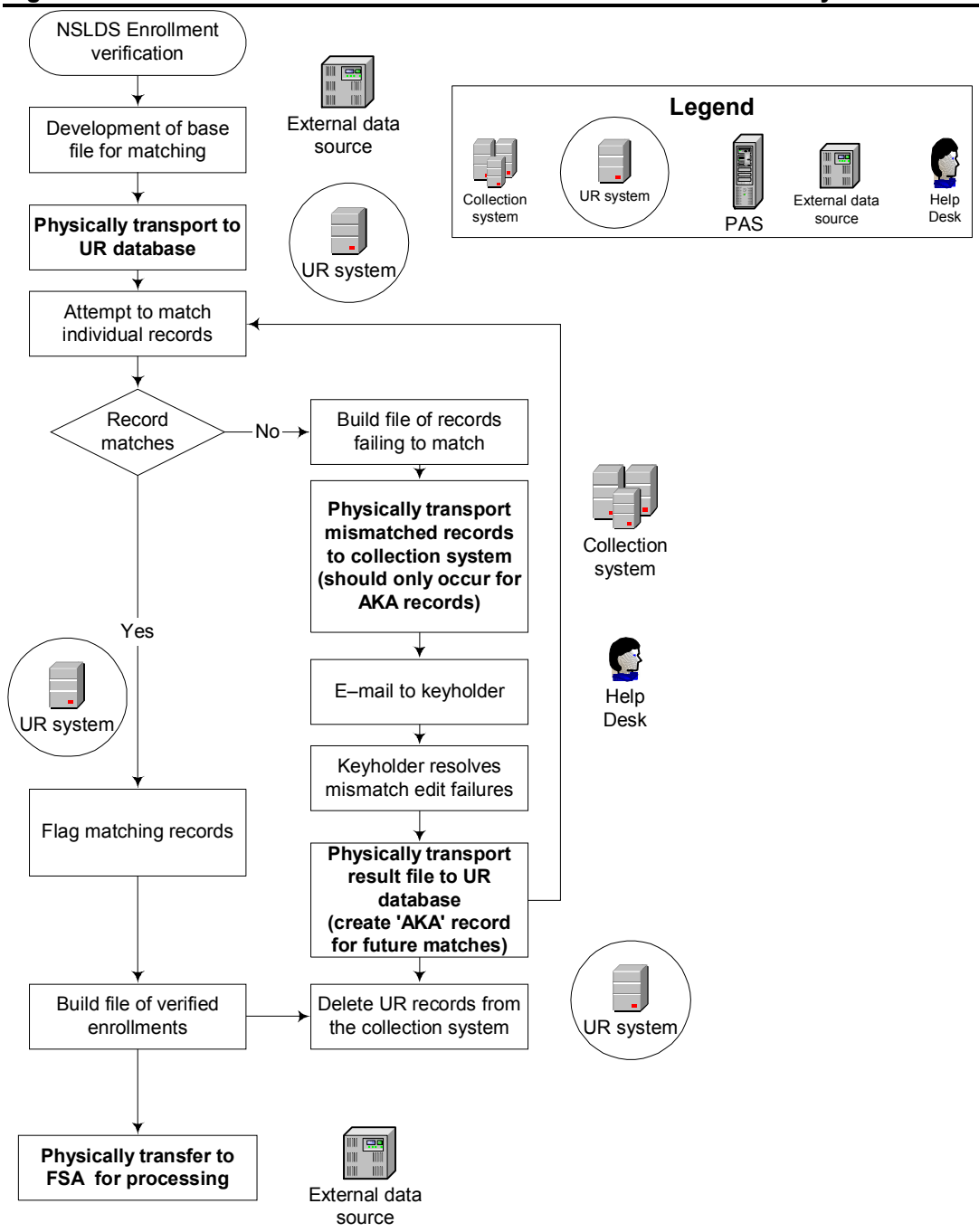
One of the proposed redisclosures—the redisclosure of mismatches of student records back to the keyholder for follow up—would be necessary for the accurate functioning of a UR system. As mentioned above, flagged student records would be sent to the keyholder after data submission, and the keyholder would work with the Help Desk to resolve the mismatches before the data were locked.

The second proposed redisclosure would involve periodic enrollment verification to NSLDS (figure 5), for which students receiving federal student loans have already consented.[44] After each collection of enrollment data passed internal and external edits and was transferred to the UR permanent storage database, a database coming from an external FSA data source would be physically transported to the UR system, where there would be an attempt to match records from the external database to the URs. For those records that matched, flags would be created and a file of verified enrollments would be built. For records that do not match, the data would be moved to the collection system where the IPEDS keyholder would be notified by email and asked to resolve the mismatches and edit failures. After these were resolved, the data would be replaced in the UR system. The records with matches would be flagged and an enrollment verification file would be built for these students. Enrollment verification files would be physically transferred to FSA for processing and the FSA records in the UR collection system would be deleted.

The redisclosure of subsequent enrollment information back to the IPEDS keyholder would provide a benefit to institutions to help mitigate the burden of submitting UR data; institutions would be "getting something back" from the process besides mandated compliance. This redisclosure would allow institutions to gain information about students who leave the institution.

[44] The proposal is for enrollment verification once per term, which OPE has said will be acceptable. Some institutions may want to report enrollment more often, and they would be allowed to do so.

Figure 5. Enrollment verification for the National Student Loan Data System



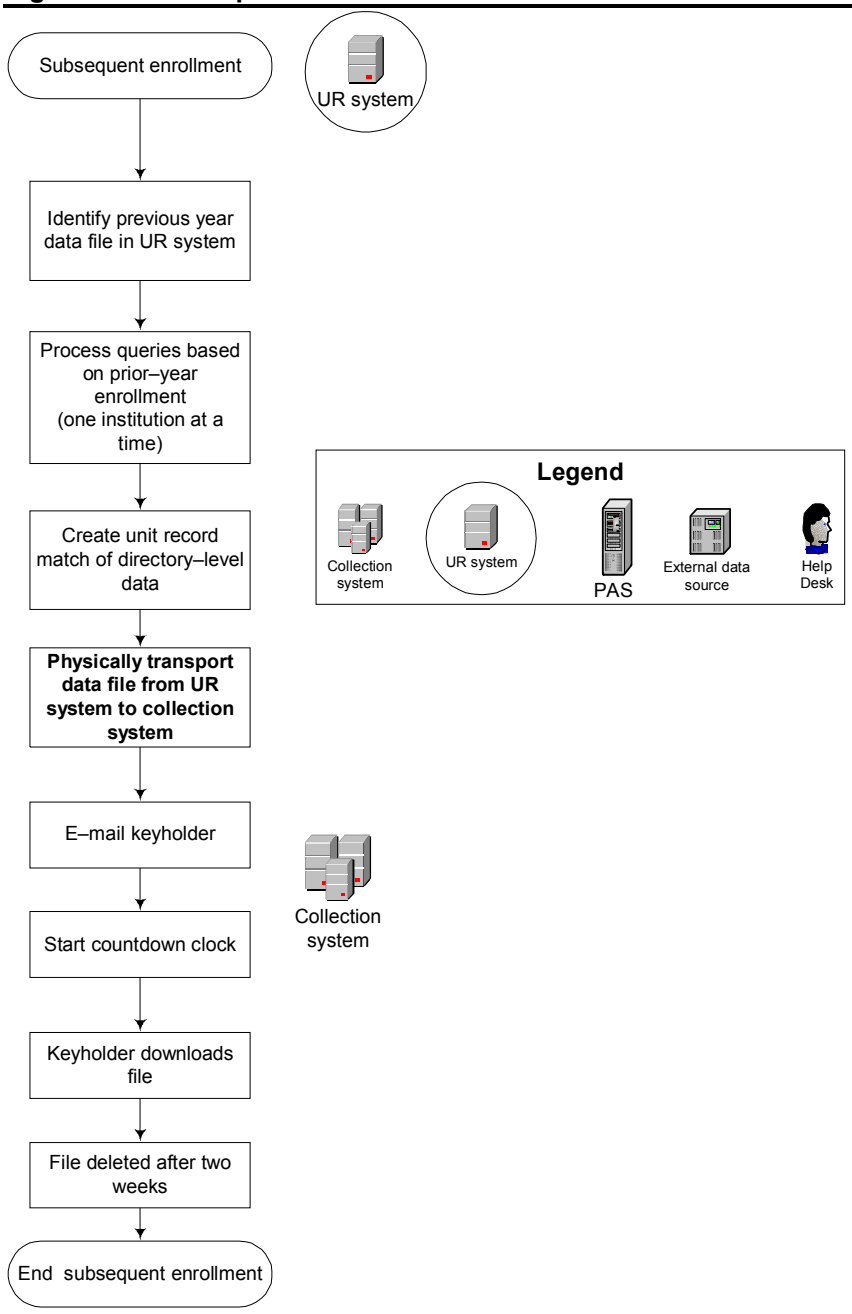
NOTE: Flowchart is for a proposed UR system; if such a system was authorized and funded, the details of the system architecture would be decided after several Technical Review Panels as well as a field test. The collection system icon refers to the IPEDS collection system for aggregate and unit record data collection. The UR system icon refers to the IPEDS offline database unit record data storage and analysis. The PAS icon refers to the IPEDS Peer Analysis System storage and analysis. The external data source icon refers to nonspecific offline data systems. The Help Desk icon refers to the IPEDS Help Desk. AKA means 'Also known as'.

One possible conceptualization of these subsequent enrollment redisclosures is presented in figure 6. The redisclosure would be conducted once a year using the previous year's enrollment and completions data. NCES would identify an institution's previous year enrollment data file in the UR database. For each institution, queries would be run for the previous year's students to check on subsequent enrollment activity. Directory level information would be compiled from the UR system, including basic information about persistence and completion during this one–year period[45]— including the subsequent school UNITID, the students' enrollment status, and the date and type of any award. The data would be physically transported from the UR system to the collection system. IPEDS keyholders would be notified by email that a data file of subsequent enrollment information for their students would be available for a limited period of time to download. A system would be developed for keyholders to access the redisclosed data in a secure environment so that they can then match them to internal records, another reason for ensuring that the matching process is as successful as possible. The file would be deleted after two weeks from the collection system. The exact process would need to be designed at a future TRP meeting, if a UR system were authorized and implemented.[46]

The data would be redisclosed back to the original institution only if each student's record does not indicate that she/he has opted out of the redisclosure. If the student has opted out, then there would be a flag documenting this response, with no additional data.[47] Given privacy concerns, institutions would be required to inform students about the use of these data for this purpose and establish a campus–wide mechanism for students to officially opt out of the redisclosure if they choose to do so. This data element would probably need to be stored as part of the header record, which could be overwritten with future submissions.

[45] Redisclosure of student information to the original institutions could take place over a longer time period if this was decided by a future design TRP and NCES.
[46] This process would include any potential restrictions to the use of redisclosed data by institutions and state coordinators.
[47] Note that institutions would still benefit from aggregate estimates generated through the UR system (for example, the percentage of student who earned a degree at a subsequent institution) that included all students, and measured what happened to students over longer periods of time.

Figure 6. Subsequent enrollment disclosure to institutions

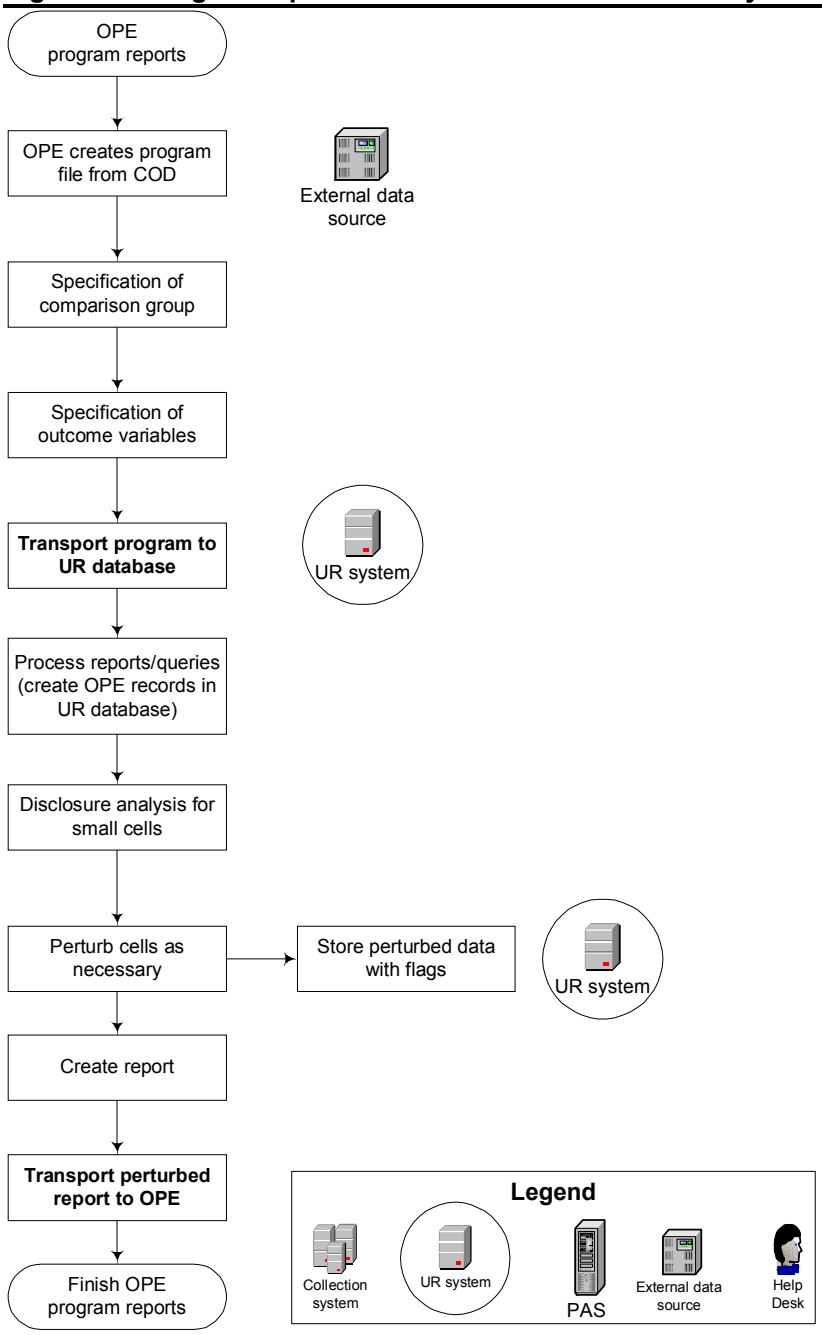


NOTE: Flowchart is for a proposed UR system; if such a system was authorized and funded, the details of the system architecture would be decided after several Technical Review Panels as well as a field test. The collection system icon refers to the IPEDS collection system for aggregate and unit record data collection. The UR system icon refers to the IPEDS offline database unit record data storage and analysis. The PAS icon refers to the IPEDS Peer Analysis System storage and analysis. The external data source icon refers to nonspecific offline data systems. The Help Desk icon refers to the IPEDS Help Desk.

Other uses of a UR system would include the generation of aggregate program reports for the Office of Postsecondary Education (OPE) in order to assess the success of student financial aid programs (figure 7). OPE staff would create a program file from the COD external data source for aid recipients and specify the comparison groups and outcome variables it wished to have reported. After NCES received the records, it would physically transport the external data file to the secure UR storage database, where query reports would be processed to match the records. NCES would create program reports with aggregate measures, after ensuring that cell sizes were large enough so that no individual could be identified and if necessary perturbing the data. The reports would be physically transported from the UR server to OPE for further analysis and dissemination. A similar process could be performed for other OPE programs.

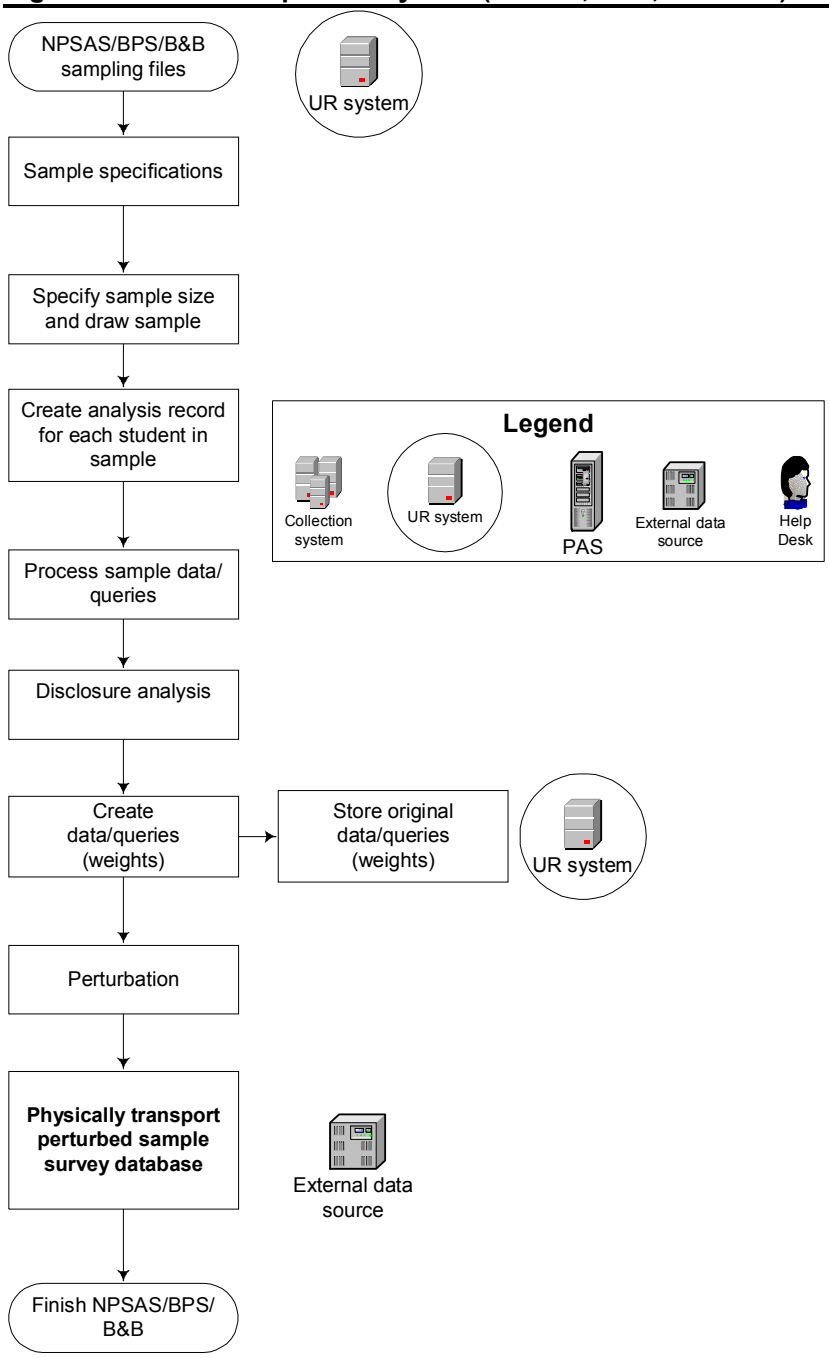
Samples of students for NCES sample surveys, such as NPSAS and BPS, could be drawn directly from the UR database. The current process for sample surveys involves drawing a sample of institutions, which are required to send UR data back for their students. The sample surveys would continue to survey students directly and merging in data from other ED databases. Figure 8 documents the process of creating NPSAS, BPS, and Baccalaureate and Beyond (B&B) sample files from the UR system. A sample size would be specified and a sample drawn from the UR system. After drawing the sample, perturbing data, and data file, the file would be appended to the sample survey database where the survey would be completed. Students would still have the possibility of opting out of NCES sample surveys, as they do now.

Figure 7. Program reports for the Office of Postsecondary Education



NOTE: Flowchart is for a proposed UR system; if such a system was authorized and funded, the details of the system architecture would be decided after several Technical Review Panels as well as a field test. The collection system icon refers to the IPEDS collection system for aggregate and unit record data collection. The UR system icon refers to the IPEDS offline database unit record data storage and analysis. The PAS icon refers to the IPEDS Peer Analysis System storage and analysis. The external data source icon refers to nonspecific offline data systems. The Help Desk icon refers to the IPEDS Help Desk. COD refers to Central Processing System.

Figure 8. NCES sample survey files (NPSAS, BPS, and B&B)



NOTE: Flowchart is for a proposed UR system; if such a system was authorized and funded, the details of the system architecture would be decided after several Technical Review Panels as well as a field test. The collection system icon refers to the IPEDS collection system for aggregate and unit record data collection. The UR system icon refers to the IPEDS off–line database unit record data storage and analysis. The PAS icon refers to the IPEDS Peer Analysis System storage and analysis. The external data source icon refers to nonspecific offline data systems. The Help Desk icon refers to the IPEDS Help Desk.

Training

Training of IPEDS keyholders and coordinators is important to ensure that the data are submitted correctly and meet the parameters of the data collection. Currently, numerous training opportunities exist in the postsecondary education community to help practitioners learn the skills of using and merging datasets. For example, the Association for Institutional Research (AIR) conducts "train the trainer" sessions generally as part of state and regional conferences, which allow institution staff to attend one close to their institution.[48] In addition, web–based tutorials are currently being developed to provide instruction for gathering data for IPEDS reporting, entering data into the data collection system, using the Peer Analysis System to produce data for analysis, and other functions to assist data providers and users. The American Association of Collegiate Registrars and Admissions Officers (AACRAO) also hosts IPEDS training sessions to inform members of changes to IPEDS and related issues, and EDUCAUSE conducts IT training. The need for these types of training would increase substantially if a UR system were to be implemented. Training would be extensive and involve multiple levels of institutional staff as well as both web–based and in–person delivery models. Other types of institutional staff, such as information technology professionals, may require training as they may become more involved in the process of IPEDS data submission at the institutional level.

Help Desk

Under the current IPEDS collection system, institutional keyholders can contact the IPEDS Help Desk to help resolve questions about data definitions, procedures, and technical problems. The Help Desk would become even more important should a UR system be implemented. Help Desk staff would be trained to answer questions from keyholders on various aspects of the data collection process. In addition, the Help Desk would ask keyholders to resolve mismatches in the URs, working closely with the keyholders to assist in this process. The proposed UR system envisions having the Help

[48] These sessions are funded through a subcontract with RTI International, which has a contract with NCES for this and other purposes.

Desk open throughout the year, whereas currently it is open around the dates of the three collection periods.

Software

The existing system for IPEDS collection and dissemination is maintained in a Microsoft Windows software environment that incorporates Active Server Page (ASP) scripting, the Visual Basic scripting language, and Microsoft SQL Server. In planning for the possibility of UR reporting, a number of requirements need to be addressed for software and hardware. Specifically, the UR collection software would need an adequate database management package capable of holding millions of student records per year; security protocols that include Secure Socket Layer (SSL), digital certificates, and password protection; and load balancing software that utilizes multiple web and database servers to prevent overload and effectively and efficiently handle large data transactions. These needs would come in addition to the current PAS environment, which would continue to expand with the addition of more years of IPEDS aggregate data files and new data elements.

Hardware

For one year of UR data, there would theoretically be four enrollment files, one completion file, and one financial aid file for each of the estimated 6,700 Title IV institutions. Disk storage space is posited to be approximately 50 GB for the first year. This includes room for the historical GRS data files that are needed and header files for all students currently in the system. The header file would remain at an estimated 4 GB per year with additional new students, with the number of term records growing slightly each year for a total of approximately 270 GB of storage needed by the end of seven years. In order to accommodate this growth, two database servers would be needed, each with 300 GB of capacity. The first server would be used for collecting the data, the second for securely storing the URs away from the Internet in final form.

The existing IPEDS collection database server does include load balancing. It might be necessary to split the UR database across two SQL Servers. The IPEDS

collection system for the five non–unit record components (S, SA, EAP, Finance, and IC) would be given its own database server. It has also been suggested that a separate database server be installed to house the IPEDS PAS, which grows annually at more than 1.5 Gigabytes (GB) per year, and the NCES DASOL applications (which include the postsecondary sample surveys and the IPEDS DAS). While the scalability of Microsoft SQL Server for handling hundreds of millions of records in the future is worthy of discussion, numerous examples of similar enterprise–wide solutions are in place in this environment across the country.

For the database server that is currently used for the IPEDS collection, the collection developer has estimated future growth of 1.5 GB per year. With new reports and tables that would be exported to other agencies for redisclosure, approximately 50 GB of disk space would be needed on this server. The PAS database server would continue to grow, with the accretion of larger and more complex files, requiring approximately 100 GB over the next seven years.

Overall, the implementation of IPEDS URs could require four dedicated database servers—one for collecting and another for storing the data separately; one for the continued collection of the nonstudent IPEDS components; and a fourth for dissemination of IPEDS via the PAS and the DASOL.

Chapter 5—Conclusions

This report has examined the feasibility of implementing a student UR system to replace the student–related components of IPEDS. As part of the feasibility study, an architecture and flow of operations for a proposed UR system, as well as a list of potential data elements that might be collected under such a system, were developed and described. In addition, the feasibility study solicited input from states and state systems; private associations of colleges and universities; institutional researchers, registrars, and financial aid officers; and other stakeholders such as the postsecondary education association community and federal agencies.

As this report has outlined, a central question for a UR system is *"Could it be done?"* The answer to this question is essentially technical. Have the information technologies and infrastructures at the campus and state levels matured, could the current IPEDS web–based reporting system be adapted to UR, and would there be adequate technical and legal protections in place at IES/NCES? The report has addressed some of the technical and system problems associated with the design and development of a new IPEDS UR system. At the technical level, a UR system could be done at most institutions given time for implementation, and the problems associated with development of such a system are manageable.

The feasibility study also addressed the "*Should it be done?"* question, providing a framework for the discussion of issues inherent in this question. These issues revolve around several areas of concern, which would need to be addressed and resolved in the design phase of a UR system should policymakers decide to authorize and fund a UR system.

Privacy is the first and more fundamental area of concern. Does the federal interest in collecting better data "trump" the right of students to control information about their enrollment, attainment, and financial aid? The confidentiality of student data would be protected to the extent allowable in the legislation under which IES/NCES operate.

Second, there would be costs and burdens to institutions associated with implementation, especially in the initial years. However, over $80 billion in federal

student financial aid presently flows to postsecondary institutions. A decision would need to be reached as to whether these direct federal benefits to institutions are sufficient to counterbalance short–term concerns about cost and the burden of implementation, or whether additional funds are needed.

Third, a UR system would require institution–level coordination, involving the cooperation of registrars, institutional research, IR, and financial aid offices on campuses; with the need to assign or perhaps hire staff. A UR system might bring fundamental changes to state coordination structures and the management of the data flow on enrollment, completions, and student aid.

A UR system would also involve issues with technological capacity and the timing of data collection and implementation. Although changes in technology could be daunting for some institutions, mechanisms would exist to help institutions with reporting. The operational timing of data collection could pose complications, but would be addressed during the design phase of UR implementation with input from institutions.

This feasibility study has outlined areas of federal interest: better information for informed consumer decisions, including the improved calculation of net prices; and more accurate measures for institutional accountability and program effectiveness, including enrollment, persistence, transfer, and attainment rates by program of study. Policymakers would be able to monitor in real–time federal student aid programs (such as Pell Grants) and variations in aid packaging. The study also has attempted to highlight some potential benefits to institutions, states, systems, consumers, and other users of NCES data.

The study did not attempt to address every challenge or make recommendations about how each aspect should be addressed. Nor did the report document specific organizational positions regarding the obstacles a UR system might face. Rather, it provided a framework for policymakers to understand the potential costs and benefits of a UR system as they discuss whether it should be considered.

The central defining question of the feasibility of a UR system in IPEDS is not a "could" question. It is a "should" question, asking whether the federal government should develop a system that is based upon individually identifiable information about enrollment, financial aid, and attainment. This system would, for the first time, give policymakers and consumers much more accurate and comprehensive information about

postsecondary education in this country. Some of the benefits of UR include the collection of new data that would measure the success rates of students at institutions to which family and federal student aid monies flow, provide more accurate consumer guidance, and improve federal programs that support those families and students. In addition to benefits, the feasibility study found a number of significant issues that would need to be overcome before a UR system could be implemented, including objections about student privacy, confidentiality of data, new institutional burdens, coordination within and outside of institutions, and timing issues. Whether a UR system should be authorized, appropriated, and implemented is left to policymakers to decide, given the benefits and constraints examined in this study.

References

Association for Institutional Research (AIR). (2000). *AIR Alert: Student Right–to–Know Act and IPEDS Graduation Rate Survey*. November. Available online at: http://www.airweb.org/page.asp?page=118.

_____. (2004a). *AIR Alert #26: IPEDS Student Unit Record Data Feasibility Study*. October. Available online at: http://www.airweb.org/page.asp?page=670.

_____. (2004b). *AIR Alert #26 Update 1: Proposed IPEDS Student Unit Record Data Feasibility Study*. November. Available online at: http://www.airweb.org/page.asp?page=711.

Berkner, L., He,S., and Cataldi, E.F. (2002). *Student Financing of Undergraduate Education: 1999–2000* (NCES 2002–167). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Carey, K. (2004 May). *A Matter of Degrees: Improving Graduation Rates in Four–Year Colleges and Universities*. Washington, DC: The Education Trust.

Choy, S. (2002). *Findings from the Condition of Education 2002: Nontraditional Undergraduates*. (NCES 2002–012). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

College Board. (2004). *Trends in College Prices 2004*. Washington, DC: College Board.

Cunningham, A. F. and Merisotis, J.P. (forthcoming). *Changes in Patterns of Prices and Financial Aid* (NCES 2005–153). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Ewell, P., Schild, P.R., and Paulson, K. (2003). *Following the Mobile Student: Can We Develop the Capacity for a Comprehensive Database to Assess Student Progression?* New Agenda Series. Bloomington, IN: Lumina Foundation for Education.

Horn, L., Chang–Wei, C., and Berker, A.. (2002). *What Students Pay for College: Changes in Net Price of College Attendance Between 1999–2000* (NCES 2002–174). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

JBL Associates. (2004). *JBL Associates Achieving the Dream Data Collection Project: Reporting System Requirements*. Unpublished document, October 6.

L'Orange, H. (2004). E–mail survey of state SHEEO and system offices in the SHEEO/NCES Network, conducted December 17.

National Center for Education Statistics. (2002a). *Statistical Standards*. Available online at: http://nces.ed.gov/statprog/2002/stdtoc.asp.

_____. (2002b). *Classification of Instructional Programs: 2000 Edition.* (NCES 2002–165.) U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

National Postsecondary Education Cooperative (NPEC). (2001 April). *Student Transition Data Systems. Report of the NPEC Working Group.* Washington, DC.

_____. (1998). *Unit Record Versus Aggregate Data: Perspectives on Postsecondary Education Data Collection, Retention, and Release. Final Report of the Working Group on Unit Record Data Versus Aggregate Data.* Washington, DC: NPEC, April.

Peng, S. (1999). *Integrated Postsecondary Education Data System (IPEDS): An improved system. Final Report of the NCES Taskforce for IPEDS Redesign*. U.S. Department of Education, Washington, DC: National Center for Education Statistics.

Pfieffer, J. (2003). *Overview of Federal Laws Pertaining to Sharing Individually Identifiable Educations Records*. Tallahassee, FL, May 22: Florida Department of Education.

Seastrom, M. (2004). *Confidentiality and Data Access at NCES*. Unpublished document.

Voorhees, R. and Barnes, G. (2003). *Data Systems to Enhance Teacher Quality*. Denver, CO: State Higher Education Executive Officers (SHEEO).