

Chapter 11

Using IPEDS for Panel Analyses: Core Concepts, Data Challenges, and Empirical Applications

Ozan Jaquette and Edna E. Parra

Introduction

Higher education researchers are increasingly using causal inference methods to answer research questions that inform higher education policy (Zhang 2010). In particular, panel modeling methods have become popular due to the growing availability of panel data and because these methods allow researchers to make inferences about causal effects that cannot be made using cross-sectional models (Cheslock and Rios-Aguilar 2011). For example, Zhang and Ness (2010) used a difference-in-difference estimation strategy to evaluate the effect of state merit aid on the out-migration of state residents. Jaquette and Curs (2013) used a panel instrumental variables estimation strategy to analyze whether declines in state financial support caused nonresident enrollments to increase at public universities.

Panel models require panel datasets, which have one observation per time period (e.g., day, year) for each unit of analysis (e.g., person, organization, state). The Integrated Postsecondary Education Data System (IPEDS) is the primary source of annual panel data on accredited postsecondary institutions that are eligible for Title IV financial aid. IPEDS and its predecessor the Higher Education General Information Survey (HEGIS) consist of survey components that cover the different aspects of postsecondary institutions (e.g., Institutional Characteristics, Finance, Completions, Fall Enrollments). We use the term postsecondary “institutions” rather than postsecondary “organizations” because the term institution has a particular

O. Jaquette, Ph.D. (✉)

Educational Policy Studies & Practice, College of Education, University of Arizona,
1430 E. Second Street, Room 327A, Tucson, AZ 85721, USA
e-mail: ozanj@email.arizona.edu

E.E. Parra

Educational Policy Studies & Practice, College of Education, University of Arizona,
1430 E. Second Street, Room 3271B, Tucson, AZ 85721, USA
e-mail: ednap@email.arizona.edu

meaning in IPEDS data collections.¹ Other sources of institution-level data, such as data about financial aid disbursements collected by the Office of Federal Student Aid, also offer possibilities for panel analyses.

This chapter provides guidance on the appropriate construction of panel analysis datasets, in particular how to construct a panel using IPEDS data, so that researchers using panel models can make the correct inferences. Research questions addressed using IPEDS data often require the creation of a panel that incorporates data from multiple HEGIS/IPEDS survey components (e.g., Finance, Completions, Institutional Characteristics) and other sources of institution-level and state-level data. Merging data from multiple survey components is a complicated process because some survey components (e.g., Completions) may measure data at the campus level but other survey components (e.g., Finance) may measure data at the institution level or the system level. Failing to consider how data complexities and limitations affect specific analyses may result in biased results and misguided policy recommendations. Therefore, the fundamental goal of this chapter is to discuss core concepts and data challenges associated with using IPEDS data to construct panel datasets that are appropriate for answering the research question at hand.

We have learned about these core concepts and data challenges through the process of constructing panel datasets to answer specific research questions. For example, we created a panel dataset from multiple HEGIS/IPEDS survey components to analyze why many liberal arts colleges became comprehensive universities from 1969–1970 to 2009–2010 (Jaquette 2013a) and to analyze the production of master's degrees by US universities from 1969–1970 to 2010–2011 (Jaquette 2013b). We also created a panel dataset from multiple IPEDS survey components, data on Pell Grant recipients, and state-level data to analyze the effect of nonresident enrollment growth on the racial and socioeconomic diversity of public universities (Curs and Jaquette 2013).

The first section of this chapter provides an introduction to institution-level data and documentation. The primary focus is on IPEDS. We also discuss HEGIS data, data from the Office of Federal Student Aid, and premade panel datasets (e.g., Delta Cost Project Database). We do not present a detailed discussion of specific survey items. Researchers interested in a change over time in specific IPEDS survey items by survey component should consult the excellent work of Fuller (2011).

The second section discusses change over time in the HEGIS/IPEDS sampling universe and the unit of analysis represented by each observation of data (e.g., campus, institution, system). Seemingly idiosyncratic data patterns are really systematic once one understands the principles behind the sampling universe and the way data are collected from multicampus institutions. Understanding these data collection principles will help researchers create panel datasets in less time and with less measurement error. Furthermore, many of these data collection principles are

¹ Institutions is a term formerly used in HEGIS and IPEDS to define an institution that was accredited at the college level by an agency or association recognized by the US Department of Education. These schools offered at least a one-year program of study creditable toward a degree, and they were eligible for participation in Title IV federal financial aid programs.

intimately related to requirements for Title IV financial aid eligibility, so readers will learn about federal financial aid policy while they learn about dataset construction.

The third section discusses “parent-child reporting,” which is the core concept of this chapter because it affects nearly all empirical applications of IPEDS panel data. Parent-child reporting occurs when multicampus institutions respond to some IPEDS surveys at the (parent) institution level and some IPEDS surveys at the (child) campus level. Parent-child reporting also occurs when distinct postsecondary institutions report some variables at the system level (e.g., a community college district office or state system office) (U.S. Department of Education. NCES 2013d). Parent-child reporting makes it difficult to merge data from different IPEDS survey components. We identify patterns of parent-child relationships and discuss solutions to challenges associated with parent-child reporting. We demonstrate how the treatment of parent-child reporting by the Delta Cost Project (DCP) database makes its use inappropriate for answering some research questions.

The fourth section utilizes the lessons from parent-child reporting to discuss solutions to common data challenges. Specifically, we discuss merging issues, appending HEGIS and IPEDS data, collapsing IPEDS data to the state level. We also discuss several changes to IPEDS that affect many empirical applications of IPEDS data, specifically change in accounting standards used in the Finance component, change in degree classification systems used by the Completions component, and change race/ethnicity categories that affect several IPEDS survey components.

Two points merit mention before we proceed. First, this chapter presents several tables based on HEGIS/IPEDS panel datasets that we have constructed. The statistics presented in these tables may differ slightly from tables presented in published reports due to data manipulation decisions we made during the process of constructing the panel datasets. Second, this chapter measures time in academic years (e.g., 2010–2011 academic year). Further, academic years refer to the data period covered by a particular survey component rather than the year the data were collected. For example, 2010–2011 IPEDS Finance data refers to institutional finances for the period 7/1/2010–6/30/2011, which corresponds to the 2010–2011 academic year, but these data were collected in spring 2012.

Access to Institution-Level Data and Documentation

After choosing a research question that requires panel analyses, researchers must construct a panel dataset to address the research question. These analysis datasets often incorporate several sources of organization-level data and also state-level or local-level data. This section provides access to data and documentation for several sources of organization-level data. Specifically, we discuss HEGIS, IPEDS, organization-level data from the Office of Federal Student Aid, and premade panel datasets that use HEGIS and/or IPEDS data (e.g., the DCP Database).

Changes over time in specific variables and variable definitions affect what analysis variables can be created for what time periods and what institutions. Readers

interested in the temporal change in IPEDS-specific survey items answered by different types of institutions should consult Fuller (2011). Readers interested in a broader discussion of organization-level data sources should consult Brint (2002).

HEGIS Data

The National Center for Education Statistics (NCES) conducted HEGIS from 1965–1966 to 1986–1987 (Fuller 2011). HEGIS was comprised of several survey components, including the Earned Degrees/Completions, Institutional Characteristics (IC), Finance, and Fall Enrollment components, all of which were conducted annually. Additional survey components, such as Residence and, Employees, Salaries, and Libraries, were conducted sporadically and, therefore, are less useful for panel analyses.

The Inter-university Consortium for Political and Social Research Data Archive located at the University of Michigan maintains the most complete archive of HEGIS data.² HEGIS data can also be downloaded from the Cornell Institute for Social and Economic Research Data Archive located at Cornell University. The Consortium for Political and Social Research website contains downloadable data and data documentation (including file layout, survey instrument, and data notes) for each available survey component and year (e.g., 1977–1978 Finance).³

IPEDS Data

Beginning with the Institutional Characteristics (IC) component, IPEDS was phased-in from 1985–1986 to 1988–1989 as a replacement for HEGIS (Fuller 2011). Currently, IPEDS collects data on nine survey components over three collection periods (fall, winter, and spring), as shown in Table 11.1 (U.S. Department of Education. NCES 2013c). The fall collection consists of the IC, Completions, and 12-Month Enrollment survey components. The winter collection consists of Student Financial Aid. The spring collection consists of Fall Enrollment, Finance, Graduation Rates, 200 % Graduation Rates, and Human Resources.

IPEDS data can be downloaded from the IPEDS Data Center, which organizes data by year and survey component.⁴ For most survey components, multiple data files exist for each year. For example, five data files are associated with the 2012

² Students and employees affiliated with organizations that are members of ICPSR can download ICPSR data at <http://www.icpsr.umich.edu/icpsrweb/content/membership/index.html>

³ To date, ICPSR has not completed the process of checking data and creating datasets in different programming languages (e.g., SAS, SPSS, and Stata). However, the raw text files and survey instruments are available for all years.

⁴ <http://nces.ed.gov/ipeds/datacenter/>

Table 11.1 Data collection cycle for IPEDS survey components labeled as “2011” on IPEDS Data Center

Survey component	Collection	Year on IPEDS Data Center	When data collected	Period covered by data	Academic year ^a	IC response flags dataset
Institutional Characteristics (IC)	Fall	2011	Fall 2011	Fall 2011 (snapshot)	2011–2012	Flags 2011
Completions	Fall	2011	Fall 2011	7/1/2010–6/30/2011	2010–2011	Flags 2011
12-Month Enrollment	Fall	2011	Fall 2011	7/1/2010–6/30/2011	2010–2011	Flags 2011
Student Financial Aid	Winter	2011	Winter 2012	7/1/2010–6/30/2011	2010–2011	Flags 2011
Fall Enrollment	Spring	2011	Spring 2012	Fall 2011 (snapshot)	2011–2012	Flags 2011
Finance	Spring	2011	Spring 2012	7/1/2010–6/30/2011	2010–2011	Flags 2011
(150 %) Graduation Rates	Spring	2011	Spring 2012	Graduation rate as of 8/31/2011 ^b	2010–2011	Flags 2011
200 % Graduation Rates	Spring	2011	Spring 2012	Graduation rate as of 8/31/2011 ^c	2010–2011	Flags 2011
Human Resources	Spring	2011	Spring 2012	November 1, 2011 (snapshot)	2011–2012	Flags 2011

^a Throughout this chapter we refer to academic year covered by the data, rather than year data collected unless otherwise specified

^b Graduation rate as of 8/31/2011 for BA seeking students who first enrolled in 2005–2006 academic year or students seeking a 2-year degree program or less who first enrolled in 2008–2009 academic year

^c Graduation rate as of 8/31/2011 for BA seeking students who first enrolled in 2003–2004 academic year or students seeking a 2-year degree or less who first enrolled in 2007–2008 academic year

IPEDS IC component, including “directory information,” “response status by survey component,” “student charges for academic year programs,” “student charges by program,” and “educational offerings, organization, admissions, services, and athletic associations.”

Documentation about IPEDS can be found within the IPEDS website.⁵ The online IPEDS Glossary provides official definitions for terms used in IPEDS data collection (e.g., what is a “Title IV institution?”).⁶ The IPEDS Resource Center provides general information about each survey component (including data collection cycle) and statutory requirements for IPEDS reporting.⁷ The Archived Survey Library (within the IPEDS Resource Center) maintains IPEDS survey forms for each survey component from 1994–1995 to present.⁸ Note that the timing of data collections is not necessarily the same as the academic year associated with the data. The Archived Survey Library also provides information on changes to IPEDS starting with data collected in 2003–2004. For example, IPEDS data collected in 2012–2013 included additional measures of distance education on the IC, Completions, and Fall Enrollments components. Finally, IPEDS has dedicated staff for each survey component; the IPEDS website lists staff by survey component and specialty/function.⁹ We have found IPEDS staff to be extremely knowledgeable, professional, and eager to answer complicated questions about the data and documentation.

Researchers must often create panel analysis variables using multiple years of data from a survey component. For each data file on the IPEDS Data Center, researchers can download a “data dictionary” that contains an overview of the data file and a list of variable names, variable labels, and variable descriptions. Researchers can triangulate information from multiple years of data dictionaries and survey forms to understand how long a variable has been collected, whether variable definitions have changed over time, and what types of institutions have non-missing values for the variable.

IPEDS panel datasets are typically created by merging panel datasets from individual survey components (e.g., Fall Enrollments and IC component from 2000–2001 to 2010–2011) by institutional identifier and academic year. Researchers must be careful when assigning academic years to each survey component. For example, IC data collected in the fall of 2012 refers to the 2012–2013 academic year, but Completions data collected in the fall of 2012 refers to the 2011–2012 academic year. Table 11.1 provides the list of IPEDS survey components that are labeled as “2011” on the IPEDS Data Center. For each survey component, Table 11.1 indicates the collection to which the survey component belongs (e.g., fall, winter, spring), the date of data collection, the time period

⁵<http://nces.ed.gov/ipeds/>

⁶<http://nces.ed.gov/ipeds/glossary/>

⁷<http://nces.ed.gov/ipeds/resource/>

⁸Note that survey forms typically differ by institutional sector (i.e., institutional control and award level).

⁹<http://nces.ed.gov/ncestaff/survetl.asp?surveyid=010>

covered, and the associated academic year (which is the time variable used in panel analyses). Table 11.1 also provides the “response flags” associated with each survey component. The response flags data file is part of the IC component and contains information—e.g., parent-child indicator variables—about each survey component (more on this later in the chapter).

While researchers typically desire an IPEDS panel analysis datasets with one observation per institution-year, the individual data files located in the IPEDS Data Center often do not have this data structure. For example, the IPEDS Data Center contains a Fall Enrollment data file called “race/ethnicity, gender, attendance status, and level of student.” This data file has one observation for each combination of institution, attendance status (full time vs. part time), and level of study (e.g., first time, first year, degree seeking; total undergraduates; graduate) and variables for each combination of race and gender. The 2011–2012 data file has 27 observations for the University of Arizona. The first observation represents the number of full-time, degree-seeking freshman, the second observation represents the number of full-time, degree-seeking transfer students, the fourth observation represents the number of full-time degree-seeking undergraduates, etc.

Therefore, creating panel datasets with one observation per institution and academic year often requires researchers to collapse data by summing down observations and/or to “reshape” data from “long” to “wide” by transforming observations (rows) into variables (columns). For example, creating the appropriate data structure from Fall Enrollment data typically requires researchers to reshape data from long to wide; instead of having observations for each combination of attendance status and level of study (e.g., part-time graduate students), there should be variables for each combination of attendance status and level of study.

Complementing IPEDS Data: The Case of Federal Student Aid Data

Timely, precise, and complete data are critical to identify problems and challenges within the higher education system. This means that researchers need to create panel datasets from all available sources of information that can deepen the understanding of issues important to the field. The different datasets available that contain information about Student Financial Aid provide a case in point. Several offices within the US Department of Education provide data about Student Financial Aid, and these data can be used to gain a broader and better understanding of the sources, types, and amounts of financial aid provided to students in our postsecondary education system. In addition, these data can be used to help understand the relationships between financial aid and policy goals promoted by the federal government and higher education institutions.

For example, two of the most common sources of Student Financial Aid information are from the Office of Postsecondary Education and the Office of Federal Student Aid. Both offices collect several sources of organization-level data on

postsecondary institutions that can be freely downloaded. Likewise, the Office of Federal Student Aid collects downloadable organization-level data on the number of disbursements and number of recipients of Title IV financial aid by program (e.g., Pell Grants, Direct Subsidized Loans).¹⁰ In addition, the Federal Student Aid Data Center provides information about federal financial aid provided to students and families through various loan and grant Title IV programs. The reports provide data about Title IV aid applicants, recipients, and disbursements. The Title IV Volume Reports are divided in three parts: (A) Loan Volume from 1999–2000 to 2012–2013, (B) Grant Volume available from 1999–2000 to 2012–2013, and (C) Campus-Based Volume data available from 2006–2007 to 2011–2012. The Office of Postsecondary Education also collects organization-level data about student loan default rates and campus crime and security statistics. The cohort default rate is a variable that in recent years has become popular because the dollar amount of student loans has been increasing over time and now surpasses all credit card debt, becoming the second largest debt in the USA. Therefore, this financial aid data can be used to complement panel datasets developed from IPEDS survey components and is available since 1992.

Faced with these multiple sources of information, a researcher may well ask, “Why should I merge information from the Office of Federal Student Aid if IPEDS has a Student Financial Aid (SFA) component?” An important reason is that the IPEDS SFA component focuses on first-time, full-time undergraduate students; therefore, if institutions have a large percentage of part-time students, financial aid disbursement by financial aid program may seem lower than institutions where the majority of students are full time. Although the financial aid report from the Office of Federal Student Aid provides information from *all* students receiving financial assistance, it does not distinguish whether the student is part time or full time, nor does it provide information about whether the student is an undergraduate or a graduate student. Also, the financial aid report from the Office of Federal Student Aid information does not have information on enrollments or other variables such as gender and sex that may be useful depending on one’s research question. Steele (2011) discussed the strengths and limitations of different sources of student financial aid data, as well as the advantages of merging from different sources.¹¹

Nonetheless, merging IPEDS data to the Office of Postsecondary Education and the Office of Financial Student Aid data can be challenging because IPEDS data utilize a different institutional identification code than the Office of Postsecondary Education and the Office of Federal Student Aid data. We discuss these merging issues in the Solutions to Common Data Challenges section later in the chapter.

¹⁰ For financial data, see <http://federalstudentaid.ed.gov/datacenter/index.html>

¹¹ <http://nces.ed.gov/pubs2012/2012834.pdf>

Premade HEGIS/IPEDS Panel Datasets

In addition to the option of creating your own panel data, several premade panel datasets using HEGIS and/or IPEDS data are available. The benefit of premade panel datasets is that researchers can avoid the time-consuming and complicated process of creating a panel from the original HEGIS or IPEDS datasets. However, premade panels also have limitations that may make them inappropriate for particular research projects.

The Delta Cost Project (DCP) database is a panel dataset that (at present) incorporates IPEDS data from 1986–1987 to 2010–2011, containing variables from the Finance, Fall Enrollments, 12-Month Enrollments, Human Resources, Graduation Rates, and Student Financial Aid survey components. The initial DCP Database was constructed by the Human Capital Research Corporation (HCRC), and it incorporated IPEDS data from 1986–1987 to 2004–2005 (Delta Cost Project 2011). HCRC provided the 2006–2008 annual updates and DCP staff provided the 2009 update. In 2012, DCP activities were split between the NCES and the Association for Institutional Research (AIR). The NCES assumed responsibility for maintaining and updating the DCP Database, and AIR assumed responsibility for the production of reports and policy briefs using the DCP.

The DCP Database has several strengths. It contains panel data on all postsecondary organizations included in IPEDS, new data are added each year, and the quality of DCP data documentation is unparalleled. The DCP Database has been used by many researchers to answer higher education policy questions. However, some researchers may be unaware that the DCP Database may be inappropriate for answering some policy questions due to the way the DCP Database handles parent-child relationships, which we will discuss later in the chapter.

A second source of premade panel data is the WebCASPAS system operated by the National Science Foundation (National Science Foundation 2013). The WebCASPAS system contains HEGIS and IPEDS data from the Completions, Fall Enrollments, Institutional Characteristics (tuition data), and Human Resource components. Unfortunately, as with the DCP Database, challenges associated with parent-child relationships make it difficult to create analysis datasets that merge WebCASPAS data with different IPEDS survey components.

A third premade dataset, the Colleges & Universities 2000 Database, spans the years 1969–1970 to 2009–2010 and contains variables from HEGIS, IPEDS, and many other sources (Brint et al. 2013). However, this database is useful for a limited range of research questions because it contains data for only 383 institutions and because measures are available at 5-year intervals (e.g., 1970, 1975).

In this section, we have outlined some of the options available for researchers interested in using the HEGIS and IPEDS in panel form. In the following section, we focus on the issue of changes over time in the HEGIS and IPEDS sampling universes and what each observation of data represents (e.g., an organization, a branch campus within an organization).

The Sample Universe and Unit of Analysis

Temporal Change in the Sample Universe

Understanding change over time in the sample universe helps researchers understand what types of organizations can be included in analyses of postsecondary organizational change. Understanding how HEGIS and IPEDS record data for multicampus organizations is also fundamental to merging data from different survey components to create an analysis dataset.

Eligibility for Title IV financial aid programs has been the primary determinant of whether an organization is included in both the HEGIS and IPEDS samples. A “Title IV institution” refers to a postsecondary institution that is eligible to enroll students that receive Title IV federal financial aid (e.g., Pell Grants, Direct Subsidized Loans) (U.S. Department of Education. NCES 2013b). A Title IV institution may be a single-campus organization or a multicampus organization. Non-Title IV institutions refer to organizations—either single campus or multicampus—that are not eligible to enroll students receiving Title IV financial aid.

HEGIS was a product of the 1965 Higher Education Act, which created Title IV financial aid programs. At that time any postsecondary institution accredited at the institutional level by an accrediting agency recognized by the Secretary of the US Department of Education was eligible to enroll students receiving Title IV financial aid (Hyatt and Dickmeyer 1980, p. 7). Although institution-level accreditation is conducted for an entire campus, “program-level” accreditation does so for a specific academic program (e.g., BS in Nursing) within a campus (Council for Higher Education Accreditation 2002). The 1965 Higher Education Act mandated that Title IV institutions would be required to complete an annual survey, leading to the creation of HEGIS (and its successor IPEDS).

The HEGIS sample universe was defined as all Title IV institutions that “offered at least a one-year program of study creditable toward a degree and they were eligible for participation in Title IV Federal financial aid programs” (U.S. Department of Education. NCES 2013b). These institutions were required to complete all HEGIS survey components.¹² Using data from the IC component, Table 11.2 shows that the number of Title IV institutions in the HEGIS sample universe increased from 2,900 in 1969 to 3,800 in 1986.

Although the contemporary IPEDS sample universe consists of all Title IV institutions, the initial IPEDS sample universe (1986–1987 to 2000–2001) included both Title IV institutions and non-Title IV institutions. IPEDS replaced not only HEGIS but also the Survey of Non-Collegiate Postsecondary Institutions and the Vocational

¹²The HEGIS sample universe included all “institutions of higher education.” An institution of higher education referred to institutions that were accredited at the institution level, were eligible for Title IV financial aid, and “offered at least a one-year program of study” (U.S. Department of Education. NCES, 2013b). What HEGIS refers to as an institution of higher education, we refer to as a Title IV institution.

Table 11.2 Number of observations in IC component, 1969–1970 to 2011–2012

Year	Frequency	Year	Frequency
1969	2,923	1991	10,445
1970	3,019	1992	10,410
1971	3,198	1993	11,005
1972	3,302	1994	10,768
1973	3,483	1995	10,608
1974	3,595	1996	10,309
1975	3,655	1997	10,180
1976	3,459	1998	9,982
1977	3,472	1999	9,822
1978	3,508	2000	9,613
1979	3,562	2001	9,564
1980	3,581	2002	7,059
1981	3,672	2003	6,998
1982	3,687	2004	7,014
1983	3,694	2005	6,939
1984	3,726	2006	6,973
1985	3,832	2007	7,006
1986	3,815	2008	7,012
1987	12,930	2009	7,125
1988	12,483	2010	7,207
1989	11,827	2011	7,393
1990	11,044	2012	7,591

Education Data System (U.S. Department of Education. NCES 1999). Referring to 1986–1987 IPEDS data, Cohen (1990) stated that “accreditation is not a requirement for inclusion in IPEDS as was the case with HEGIS: rather, accreditation is a characteristic about which data are collected in IPEDS.” Whereas the HEGIS sample universe was all Title IV institutions, the 1986–1987 IPEDS sample universe all “postsecondary institutions” which were defined as organizations “whose primary purpose is the provision of postsecondary education” (Cohen 1990, p. 511). Therefore, the 1986–1987 sample universe comprised all Title IV institutions included in HEGIS and also non-Title IV institutions that had been excluded from HEGIS. Table 11.2 indicates that the number of observations in the IC component increased from about 3,800 in the 1985–1986 academic year to about 13,000 in 1986–1987 with the inception of IPEDS.

The sector variable determined which IPEDS survey components each institution had to complete from 1986–1987 to 1999–2000 (Fuller 2011). Postsecondary institutions are categorized into nine sectors based on the dimensions of control (public, private not-for-profit, and private for-profit) and highest degree awarded (baccalaureate or higher; 2-year but less than baccalaureate; and less than 2-year). Table 11.3 displays the number of observations by sector from 1986–1987 to 2011–2012, using data from the IC component. Postsecondary institutions offering a two-year degree (e.g., associate’s) or above were defined as degree-granting institutions,

Table 11.3 Number of observations by sector, IC component 1986–1987 to 2011–2012

Award lvl		Baccalaureate or above			2-year but less than 4-year			Less than 2-year			Total
Control		Public	Non- profit	For- profit	Public	Non- profit	For- profit	Public	Non- profit	For- profit	
Sector	0 (admin)	1	2	3	4	5	6	7	8	9	
1987	170	641	1,960	144	1,267	923	892	533	563	5,769	12,862
1988	154	641	1,945	120	1,262	849	852	380	516	5,717	12,436
1989	141	633	1,907	112	1,244	773	877	321	467	5,252	11,727
1990	143	624	1,873	106	1,224	732	862	301	388	4,666	10,919
1991	123	623	1,867	114	1,221	629	845	280	361	4,224	10,287
1992	120	625	1,880	121	1,255	630	811	278	350	4,194	10,264
1993	126	626	1,976	149	1,272	627	758	277	372	4,703	10,886
1994	117	631	1,987	172	1,272	616	736	277	340	4,503	10,651
1995	105	629	2,005	185	1,284	627	746	299	335	4,293	10,508
1996	96	632	2,004	206	1,284	615	700	306	308	4,065	10,216
1997	94	635	2,002	217	1,288	599	988	278	303	3,691	10,095
1998	96	641	2,003	235	1,280	555	965	366	299	3,456	9,896
1999	91	645	1,998	264	1,269	527	970	365	303	3,312	9,744
2000	82	647	1,991	294	1,256	492	961	312	292	3,169	9,496
2001	83	651	1,991	348	1,254	459	932	361	292	3,142	9,513
2002	81	651	1,686	338	1,192	289	800	312	130	1,417	6,896
2003	80	655	1,673	312	1,181	274	788	306	132	1,422	6,823
2004	83	656	1,686	363	1,185	254	806	290	137	1,415	6,875
2005	83	661	1,648	381	1,165	236	817	272	128	1,413	6,804
2006	83	660	1,641	422	1,171	229	839	248	111	1,440	6,844
2007	84	665	1,629	465	1,164	219	859	240	104	1,462	6,891
2008	84	674	1,631	502	1,150	189	873	239	101	1,459	6,902
2009	84	674	1,629	542	1,143	195	915	236	94	1,522	7,034
2010	83	692	1,636	577	1,113	188	986	241	97	1,522	7,135
2011	81	698	1,637	662	1,098	182	1,043	254	98	1,614	7,367
2012	81	702	1,649	747	1,086	192	1,074	260	93	1,707	7,591

and postsecondary institutions whose highest degree awarded was less than 2 years were named nondegree institutions. From 1986–1987 to 1999–2000, all degree-granting institutions (sectors 1–6) and all public non-degree-granting institutions (sector 7) were required to complete all IPEDS survey components. Private non-profit non-degree-granting institutions (sector 8) and private for-profit non-degree-granting institutions (sector 9) were required to complete the IC component from 1986–1987 to 1989–1990 and the Consolidated Survey from 1990–1991 to 1999–2000, but not the remaining survey components (Fuller 2011).

Over time, Title IV eligibility became the primary criterion for inclusion in the IPEDS sample universe. Section 490 of the Higher Education Amendments of 1992 (P.L. 102–325) affirmed that completion of the IPEDS survey is mandatory for institutions participating in any Title IV program (U.S. Department of Education. NCES 2005). The US Department of Education’s NCES (1999) reported on redesigning IPEDS stated that the completion of all IPEDS survey

Table 11.4 Title IV eligibility flags 1999–00 to 2009–10, IC data

	Not eligible/missing data	Eligible	Branch campus	Deferment
2000	2,863	6,460	116	57
2001	2,733	6,613	116	51
2002	165	6,567	126	38
2003	171	6,503	105	44
2004	152	6,553	132	38
2005	138	6,509	127	30
2006	101	6,593	124	26
2007	108	6,637	120	26
2008	93	6,568	217	24
2009	154	6,652	211	17
2010	125	6,836	152	22

components would be mandatory for all Title IV institutions and mandatory for all postsecondary institutions applying for Title IV eligibility but would be optional for non-Title IV institutions. Table 11.4 indicates that the number of non-Title IV institutions included in the IC component decreased from 2,733 in the 2000–2001 academic year to 165 in 2001–2002.

To summarize, HEGIS and IPEDS data can be used to conduct panel analyses of Title IV institutions. The HEGIS sample universe consisted of all Title IV institutions that offered at least a one-year program of study, and these institutions were required to complete all IPEDS survey components. The IPEDS sample universe also included all Title IV institutions. However, from 1986–1987 to 1999–2000, Title IV institutions that were private nonprofit non-degree-granting (sector 8) and private for-profit non-degree-granting (sector 9) were not required to complete all IPEDS survey components; rather, they were required to complete either the IC component (1986–1987 to 1989–1990) or the Consolidated Survey (1990–1991 to 1999–2000) (Fuller 2011). From 1986–1987 to 2000–2001, IPEDS collected data on non-Title IV institutions. Depending on sector and year, these non-Title IV institutions were required to complete one of the following: (1) all survey components, (2) only the Institutional Characteristics, or (3) the Consolidated Survey component. Researchers interested in analyzing the behavior of non-Title IV institutions after 2000–2001 must use non-IPEDS data sources, such as data from state private postsecondary education regulatory agencies utilized by Cellini and Goldin (2012).

Potential problems with change over time in the sample universe are further complicated in the case of multicampus university systems, which are subject to Program Participation Agreements. We discuss this issue in the following section.

Program Participation Agreements

Program Participation Agreements (PPA) play an important role in IPEDS and have important implications for the creation of IPEDS panel datasets. A PPA is one of several requirements—including accreditation and legal authorization to operate in

the state—for a postsecondary institution to be considered a Title IV institution (Congressional Research Service 2007). A PPA is “a written agreement between a postsecondary institution and the Secretary of Education” that allows institutions to participate in Title IV financial aid programs and “conditions the initial and continued participation of an eligible institution in any Title IV program” (U.S. Department of Education. NCES 2013b). PPAs stipulate that the completion of all IPEDS survey components is one requirement to maintain Title IV eligibility (Legal Information Institute 2013).

PPAs apply to each branch campus of a multicampus Title IV institution (Legal Information Institute 2013). The IPEDS Glossary defines a branch institution as “a campus or site of an educational institution that is not temporary, is located in a community beyond a reasonable commuting distance from its parent institution, and offers full programs of study, not just courses” (U.S. Department of Education. NCES 2013b). Branches that do not have a separate PPA but are branches of a Title IV institution with a separate PPA are eligible to enroll students who receive Title IV financial aid. We refer to a Title IV main campus as the main campus of a multicampus Title IV institution that has its own PPA that it shares with branch campuses. We refer to a Title IV branch campus when a campus enrolls students receiving Title IV financial aid under the PPA of a Title IV institution, but the branch campus does not have its own PPA.

The example of Pennsylvania State University and the University of Wisconsin help illustrate these concepts. Pennsylvania State University has 24 campuses including the main campus at University Park. The University Park campus has a PPA that it shares with the branch campuses (e.g., Altoona, Harrisburg). Therefore, the 24 campuses of Pennsylvania State University represent a single Title IV institution with a Title IV main campus at University Park and 23 Title IV branch campuses. In contrast, the University of Wisconsin System includes 11 four-year universities (e.g., Madison, Milwaukee, Stout) (University of Wisconsin System 2013). Each four-year university in the University of Wisconsin System has their own PPA. Therefore, the 11 four-year universities in the University of Wisconsin represent 11 Title IV institutions. Whereas Pennsylvania State University–Harrisburg and University of Wisconsin–Stout may appear similar to a casual observer (e.g., both are “regional” state universities), they are different kinds of entities from the perspective of IPEDS.

For a researcher creating panel data from IPEDS, these distinctions become important when trying to merge institutions across survey components. Each Title IV institution receives an Office of Postsecondary Education ID code (OPEID). OPEID is a superior institutional identifier to UNITID—the primary institutional identifier used in IPEDS—because it is a hierarchical code that can be used to group Title IV branch campuses with a Title IV main campus. The variable OPEID has been included on the IPEDS IC survey component since 1995–1996. Table 11.5 displays the OPEID for selected campuses from Pennsylvania State University. Each unique OPEID is an eight-digit number, where the first six digits of the OPEID uniquely identify a specific Title IV institution with a Program Participation Agreement. For example, the first six digits of the OPEID for all 24 campuses of Pennsylvania State University are 003329. The two-digit suffix of OPEID identifies

Table 11.5 2010–11 OPEID and UNITID for all Pennsylvania State University, 2010–11 IC data

Institution	6-digit OPEID	OPEID suffix	8-digit OPEID	UNITID
Pennsylvania State University-Main Campus	003329	00	00332900	214777
Pennsylvania State University-College of Medicine	003329	01	00332901	214616
Pennsylvania State University-Penn State Lehigh Valley	003329	02	00332902	214670
Pennsylvania State University-Penn State Altoona	003329	03	00332903	214689
Pennsylvania State University-Penn State Beaver	003329	04	00332904	214698
Pennsylvania State University-Penn State Erie-Behrend College	003329	05	00332905	214591
Pennsylvania State University-Penn State Berks	003329	06	00332906	214704
Pennsylvania State University-Penn State Dubois	003329	07	00332907	214740
Pennsylvania State University-Penn State Fayette- Eberly Campus	003329	08	00332908	214759
Pennsylvania State University-Penn State Hazleton	003329	09	00332909	214768
Pennsylvania State University-Penn State Greater Allegheny	003329	10	00332910	214786
Pennsylvania State University-Penn State Mont Alto	003329	11	00332911	214795
Pennsylvania State University-Penn State New Kensington	003329	12	00332912	214625
Pennsylvania State University-Penn State Abington	003329	13	00332913	214801
Pennsylvania State University-Penn State Schuylkill	003329	14	00332914	214810
Pennsylvania State University-Penn State Worthington Scranton	003329	15	00332915	214652
Pennsylvania State University-Penn State Shenango	003329	16	00332916	214634
Pennsylvania State University-Penn State Wilkes-Barre	003329	17	00332917	214643
Pennsylvania State University-Penn State York	003329	18	00332918	214829
Pennsylvania State University-Penn State Great Valley	003329	19	00332919	214607
Pennsylvania State University-Penn State Harrisburg	003329	20	00332920	214713
Pennsylvania State University-Brandywine	003329	21	00332921	214731
The Dickinson School of Law of the Pennsylvania State University	003329	22	00332922	212018

Table 11.6 2010–11 OPEID and UNITID for University of Wisconsin campuses, 2010–11 IC data

Institution	6-digit OPEID	OPEID suffix	8-digit- OPEID	UNITID
University of Wisconsin-Madison	003895	00	00389500	240444
University of Wisconsin-System Administration	003895	A1	003895A1	240435
University of Wisconsin Extension	003895	A3	003895A3	260381
University of Wisconsin-Milwaukee	003896	00	00389600	240453
University of Wisconsin-Green Bay	003899	00	00389900	240277
University of Wisconsin-Stout	003915	00	00391500	240417
University of Wisconsin-Eau Claire	003917	00	00391700	240268
University of Wisconsin-La Crosse	003919	00	00391900	240329
University of Wisconsin-Oshkosh	003920	00	00392000	240365
University of Wisconsin-Platteville	003921	00	00392100	240462
University of Wisconsin-River Falls	003923	00	00392300	240471
University of Wisconsin-Stevens Point	003924	00	00392400	240480
University of Wisconsin-Superior	003925	00	00392500	240426
University of Wisconsin-Whitewater	003926	00	00392600	240189
University of Wisconsin-Parkside	005015	00	00501500	240374

each campus that shares the Program Participation Agreement. A Title IV main campus has the two-digit suffix “00” that signifies that this institution holds a Program Participation Agreement. For example, the OPEID for the Pennsylvania State University at University Park is 003329-00. A Title IV branch campus has the same six-digit OPEID as the Title IV main campus, but a different two-digit suffix. For example, the OPEID for Pennsylvania State University–Harrisburg is 003329-20. In contrast to Pennsylvania State University, the first six digits of OPEID in Table 11.6 differ from one another for each university in the University of Wisconsin System because each university in the University of Wisconsin System is a distinct Title IV institution with a separate Program Participation Agreement.

As mentioned earlier, each Title IV institution must complete all IPEDS survey components (U.S. Department of Education. NCES 2013a). For example, each four-year university in the University of Wisconsin System is a Title IV institution with a separate Program Participation Agreement and, therefore, must complete all IPEDS survey components. Multicampus Title IV institutions can decide whether they want to report IPEDS data for Title IV branch with the main campus or to report IPEDS data separately for each Title IV branch campus. For example, the University of Arizona chooses not to report separate IPEDS data for the University of Arizona South, which is a branch campus that offers BA, MA, and certification programs (UA South 2013). By contrast, Pennsylvania State University reports separate IPEDS data for each of the 24 campuses, but this is optional because all 24 campuses share the University Park campus PPA. Often Title IV branch campuses report separate IPEDS data for some survey components (e.g., Fall Enrollments) but report data for other survey component for the main campus, which is known as parent-child reporting (U.S. Department of Education. NCES 2013d). For example,

in 2004–2005 Pennsylvania State University–Harrisburg (OPEID=003329-20) reported Fall Enrollment data but did not report Finance data; the Finance data for Pennsylvania State University–University Park (OPEID=003329-00) included the data for the Harrisburg campus.

Given the above-mentioned complexities, researchers must pay very close attention to the way that multicampus institutions report data in IPEDS in order to ensure that the units of analysis are consistent. We discuss the issue of units of analysis further in the next section.

Unit of Analysis

Having discussed HEGIS/IPEDS sampling universes and Program Participation Agreements, we can discuss what each observation of HEGIS and IPEDS data represents. Panel datasets are organized by a time period (e.g., days, years) and a unit of analysis (e.g., students, organizations, campuses). The time period in HEGIS/IPEDS panel datasets is typically the academic year. Researchers often assume that organizations are the unit of analysis in HEGIS/IPEDS panel datasets; however, as shown in the previous section, each observation of HEGIS and IPEDS data does not necessarily represent an organization. Furthermore, within a single year of data, the units of observation present in a particular IPEDS survey component may differ from those in another IPEDS survey component, which makes it difficult to create a panel dataset that merges variables from multiple survey components. Therefore, understanding what kind of entity is represented by each observation of HEGIS and IPEDS data is vital to the creation of panel analysis datasets.

To reiterate the sampling universe discussion, the HEGIS and IPEDS collect data on Title IV institutions (we ignore non-Title IV institutions, which were in the IPEDS sample universe from 1986–1987 to 2000–2001). Title IV institutions may be single-campus organizations or multicampus organizations. Title IV branch campuses do not have separate institution-level accreditation, but they are eligible to enroll students receiving Title IV financial aid as branches of a main campus or system office that has institution-level accreditation. For the purpose of panel analyses, it is often useful to consider each Title IV institution as a distinct postsecondary organization, even though the difference between some Title IV institutions and Title IV branch campuses may sometimes appear semantic (e.g., University of Wisconsin–Stout vs. Pennsylvania State University–Harrisburg).

NCES refers to a “system” as two or more Title IV institutions under the control of a common governing body (U.S. Department of Education. NCES 2013b). The distinction between a system of postsecondary organizations and a multicampus postsecondary organization can often appear vague. For example, the University of Texas System (2013) has nine “universities” (e.g., Austin, San Antonio, El Paso), the University of California (2013) has 10 “campuses” (e.g., Berkeley, Irvine, Davis), and the Pennsylvania State University (2013) has 24 “campuses” (e.g., University

Park, Harrisburg). Focusing on the concept of a Title IV institution helps make the distinction between a system and a multicampus organization clearer for the purposes of data analysis. The University of Texas System is a system because each constituent “university” is a Title IV institution with separate institution-level accreditation. Likewise, the University of California is a system because each constituent “campus” is a Title IV institution with separate institution-level accreditation. In contrast, Pennsylvania State University is not a system because a system consists of two or more Title IV institutions, and Pennsylvania State University is a single Title IV institution with multiple Title IV branch campuses.

Change over time HEGIS/IPEDS data collection procedures affect whether each observation of data for a particular survey component represents a Title IV institution, a Title IV branch campus, or a system office. During HEGIS, each Title IV institution was required to complete all survey components. Title IV institutions included both single-campus institutions and multicampus institutions where branch campuses were not accredited at the institution level.¹³ Each Title IV institution was uniquely identified by the Federal Interagency Committee on Education (FICE) code. Title IV branch campuses that did not have separate institution-level accreditation did not receive a FICE code, and their data were reported with the Title IV institution.¹⁴ Therefore, each observation of HEGIS data represented either a single-campus Title IV institution or a multicampus Title IV institution (U.S. Department of Education, NCES 1976).

Similar to HEGIS, IPEDS requires each Title IV institution to complete all survey components.¹⁵ The unique institutional identifier for IPEDS is called the “UNITID” (sometimes called the IPEDS ID). Unlike HEGIS, Title IV branch campuses that choose to report IPEDS data receive UNITIDs. Prior to 2004, Title IV branch campuses that chose to report IPEDS data always reported IC data but had the option of reporting data for other survey components along with the main campus (Association for Institutional Research 2003). For example, given that the Pennsylvania State University campuses chose to report IPEDS data, the Harrisburg (UNITID=003329) campus was required to report IC data but had the option of reporting Fall Enrollment and Finance data along with the University Park campus (UNITID=214777). After 2004, Title IV branch campuses that chose to report IPEDS data were required to report separate data for all campuses except Finance. Therefore, given that Pennsylvania State University campuses chose to report

¹³ Additionally, HEGIS collected limited data (e.g., IC, some Finance data) on system offices.

¹⁴ The formal language for this rule is as follows: “Please note that each institution, branch, campus, or other entity separately certified by the accreditation and institutional eligibility unit of the U.S. Office of Education, with its own FICE code, and listed separately in the Education Director of Higher Education, should be reported on a separate survey form and not included or combined with any other such certified unit. Branches, campuses, and other organizational entities not separately certified should be included with the appropriate institution or branch report” (U.S. Department of Education, NCES 1976, p. 46).

¹⁵ However, from 1986–1987 to 1999–2000, certain Title IV institutions were not required to complete all IPEDS survey components as discussed in the Sampling Universe section.

IPEDS data, the Harrisburg campus was required to report separate data for all survey components except for Finance.

Ignoring non-Title IV institutions, each observation in IPEDS data from a particular survey component is represented by a unique UNITID code. Yet this code can represent one of five different entities: (1) a single-campus Title IV institution; (2) a multicampus Title IV institution where branch campuses do not report IPEDS data and do not have UNITIDs (e.g., University of Arizona South); (3) the main campus of a multicampus Title IV institution where the Title IV branch institutions report data separately (e.g., Fall Enrollment data in 2004–2005 for Pennsylvania State University–University Park); (4) a Title IV branch campus that reports data separately from the main campus (e.g., Fall Enrollment data in 2004–2005 for Pennsylvania State University–Harrisburg); or (5) a multicampus Title IV institution where the Title IV branch campuses report data with the main campus for that particular survey component (e.g., Finance data in 2004–2005 for Pennsylvania State University–University Park). As one might expect, this complexity presents challenges for a researcher trying to match institutions across different IPEDS survey components.

To summarize, panel datasets are defined by a measure of time (e.g., academic years) and a unit of analysis (e.g., organizations). A common unit of analysis in IPEDS panel analysis is the organization, and we argue that each Title IV institution should be considered a distinct organization for most analyses. The campus is another potential unit of analysis. Multicampus Title IV institutions are comprised of a Title IV main campus and one or more Title IV branch campuses. HEGIS did not contain observations for Title IV branch campuses, but IPEDS does. Researchers attempting to create an IPEDS panel dataset should be aware that each observation of data in IPEDS survey components does not necessarily represent a distinct organization. Rather, each observation of IPEDS data may represent a Title IV institution (which can be considered a distinct organization), a Title IV main campus (not a distinct organization), a Title IV branch campus (not a distinct organization), or a system (two or more distinct organizations).

The implications for researchers depend on their research question and whether it requires them to use data from one or more survey components. IPEDS panel analyses that use only one survey component (e.g., Fall Enrollments) may be able to use the campus as the unit of analysis for those multicampus Title IV institutions that choose to report branch campuses separately. IPEDS panel analyses that require a dataset that merges data from multiple survey components must often use Title IV institutions as the unit of analysis. Title IV branch campuses in IPEDS often report IC data but report data for other survey components with the main campus or the system office, which is called parent-child reporting. Parent-child reporting differs across IPEDS survey components and has changed over time (Association for Institutional Research 2003; U.S. Department of Education. NCES 2012). Parent-child reporting affects most panel analysis applications of IPEDS data and, in particular, makes it difficult to create a panel dataset with variables from multiple IPEDS survey components. The next section dissects parent-child reporting and discusses how to overcome challenges associated with this issue.

Parent-Child Reporting

Parent-child relationships are intimately linked to the organizational structure of an institution and the unit of analysis. A parent-child relationship occurs when data for a “child” campus (e.g., a branch) is reported by a “parent” campus, for a particular variable in a survey component. Parent-child relationships not only change over time but can also differ from one survey component to another, making construction and analysis of panel data a difficult task. For example, a child campus may report Finance variables with the parent campus, but that same child campus may report Fall Enrollment variables separately from the parent campus.

Parent-child relationships affect nearly all empirical applications of HEGIS/IPEDS panel data. Parent-child relationships make it difficult to (a) merge data from different HEGIS/IPEDS survey components, (b) merge HEGIS/IPEDS data to other sources of organization-level data, and (c) conduct analyses using a campus-level unit of analysis. This section describes parent-child relationships in detail, explains how they affect empirical research, identifies patterns of parent-child relationships across survey components and institutional types, and discusses solutions to challenges associated with the parent-child problem, with a focus on “collapsing” solutions. We explain the collapsing solution utilized by the DCP Database and use an empirical example to demonstrate how the DCP collapsing solution can lead researchers to draw erroneous conclusions.

What Is Parent-Child Reporting?

Parent-child reporting occurs when multicampus Title IV institutions report data for a Title IV branch campus as part of the Title IV main campus data: “the parent institution (normally the main campus) submits a combined report, which includes data for its own (main) campus plus data for one or more branch campuses (these branches, because they do not report data separately from the main campus or the system office, are referred to as “children”)” (U.S. Department of Education. NCES 2013d). Parent-child reporting for a multicampus Title IV institution often differs across survey components. For example, in 2004–2005, each of the 24 Pennsylvania State University campuses reported separate Fall Enrollment data, but they all reported Finance data as part of the University Park campus. Parent-child reporting also occurs when two or more Title IV institutions (either single campus or multicampus) that are part of the same system report some Finance variables at the system office level rather than reporting these variables separately for each Title IV institution.

For the remainder of this chapter, the concept “child” will refer to a UNITID-year observation that is a child in a parent-child relationship in some survey component. The concept “parent” will refer to a UNITID-year observation that is a parent in a parent-child relationship in some survey component. “Child-level”

observations will refer to UNITID observations in a parent-child relationship where there are distinct observations for the parent and for the child campuses. “Parent-level” observations will refer to UNITID observations in a parent-child relationship where observations for the child campuses do not exist and observations for the parent campus include combined data from both the parent and child campuses. Additionally, campus-level data refers to data that has distinct observations for each campus (e.g., Title IV main campus, Title IV branch campus). Organization-level data refers to data where each observation represents a Title IV institution (either single campus or multicampus).

Rutgers University provides an example of the parent-child relationship complexities. Rutgers University is a three-campus Title IV institution with a Title IV main campus (New Brunswick) and two Title IV branch campuses (Camden and Newark). Additionally, IPEDS also reported data for the Rutgers University Central Office in some years, but we would not refer to Rutgers University as a system because it consists of one Title IV institution rather than two or more Title IV institutions. Table 11.7 displays raw Fall Enrollments data for all Rutgers University campuses for selected variables over the 1986–1987 to 2010–2011 period. Note that each of the three campuses had enrollment data for each year. The variable “parid_ef” indicates the UNITID of the parent institution for the Fall Enrollment component. Parid_ef was missing for all observations from 1986–1987 to 2010–2011 because each Rutgers University campus reported Fall Enrollment data separately for each year, rather than reporting with the parent.

Table 11.8 displays raw Finance data for all Rutgers University campuses for selected variables from 1986–1987 to 2010–2011.¹⁶ In 1986–1987, 1987–1988, and 1990–1991, the Finance data contained observations for each Rutgers University campus; however, all finance variables were missing for the Camden and Newark campus observations and the observations for the New Brunswick campus contained the aggregated data from all three campuses. Therefore, the New Brunswick campus was the “parent” campus—as indicated by the variable “parid_f”—and the Camden and Newark campuses were child campuses. In 1988–1989, 1990–1991, and 2004–2005 to 2010–2011, the Finance data contained an observation labeled “New Brunswick campus,” but no observations for the Camden or Newark campuses. Here, the New Brunswick observations contained the aggregated data for all three campuses. From 1991–1992 to 2003–2004, the Finance data contained observations for the Rutgers Central Office but not for the New Brunswick, Camden, or Newark campuses; the Central Office observations contained aggregated data for all three of these campuses. In this case, the Central Office was the parent campus—as indicated by the variable “parid_f”—and the New Brunswick, Camden, and Newark campuses were child campuses. The example of Rutgers University demonstrates how parent-child relationships can vary across survey components and the designation of the parent institution can change over time.

¹⁶The variable “net assets” became available only after Rutgers transitioned from “common” accounting standards to “GASB 34/35” accounting standards in the 2001–2002 academic year.

Table 11.7 Fall Enrollments, Rutgers University, selected years, selected variables

Institution	UNITID	PARID_EF	Year	Undergrad full-time	Undergrad part-time	Grad full-time	Grad part-time
Rutgers, Camden	186371	-	1987	2,364	1,488	559	548
Rutgers, New Brunswick	186380	-	1987	22,090	3,930	2,386	5,563
Rutgers, Newark	186399	-	1987	4,089	2,102	1,067	2,353
Rutgers, Camden	186371	-	1990	2,717	1,406	672	542
Rutgers, New Brunswick	186380	-	1990	21,514	3,763	2,621	5,122
Rutgers, Newark	186399	-	1990	3,837	1,965	1,245	2,436
Rutgers, Camden	186371	-	1995	2,457	1,146	698	532
Rutgers, New Brunswick	186380	-	1995	21,281	3,619	2,789	5,775
Rutgers, Newark	186399	-	1995	4,000	1,967	1,182	2,328
Rutgers, Camden	186371	-2	2000	2,594	943	741	658
Rutgers, New Brunswick	186380	-2	2000	24,820	2,979	2,753	4,756
Rutgers, Newark	186399	-2	2000	4,059	1,717	1,178	2,267
Rutgers, Camden	186371	-2	2005	3,060	947	777	779
Rutgers, New Brunswick	186380	-2	2005	24,416	2,397	3,991	3,892
Rutgers, Newark	186399	-2	2005	5,035	1,573	1,311	2,374
Rutgers, Camden	186371	-2	2010	3,342	779	908	752
Rutgers, New Brunswick	186380	-2	2010	27,588	1,507	4,559	3,712
Rutgers, Newark	186399	-2	2010	5,754	1,553	1,719	2,475
Rutgers, Camden	186371	-2	2011	3,629	868	903	758
Rutgers, New Brunswick	186380	-2	2011	28,904	1,447	4,951	3,610
Rutgers, Newark	186399	-2	2011	5,880	1,599	1,812	2,507

PARID_EF: Identification Number for Parent Institution for Fall Enrollment component

Table 11.8 Finance, Rutgers University, selected years, selected variables

Institution	UNITID	PARID_F	Year	Account	State approp.	Current rev	Land value	Net assets
Rutgers, New Brunswick	186380	186380	1989	common	313,571,079	650,283,291	22,862,125	-
Rutgers, New Brunswick	186380	186380	1990	common	320,526,000	707,045,000	-	-
Rutgers, Camden	186371	186380	1991	common	-	-	-	-
Rutgers, New Brunswick	186380	186380	1991	common	307,768,000	735,131,000	22,991,000	-
Rutgers, Newark	186399	186380	1991	common	-	-	-	-
Rutgers, Central Office	186362	186362	1992	common	336,382,000	804,843,000	26,211,000	-
Rutgers, Central Office	186362	186362	1993	common	332,237,000	826,937,000	27,726,000	-
Rutgers, Central Office	186362	186362	1994	common	351,842,000	881,718,000	29,284,000	-
Rutgers, Central Office	186362	186362	1995	common	368,730,000	935,206,475	30,732,000	-
Rutgers, Central Office	186362	186362	1996	common	375,548,000	1,055,166,000	30,807,000	-
Rutgers, Central Office	186362	-	1997	common	373,798,000	983,445,000	33,472,000	-
Rutgers, Central Office	186362	-2	1998	common	388,237,000	1,038,362,000	33,694,000	-
Rutgers, Central Office	186362	-2	1999	common	397,072,000	1,100,557,000	34,609,000	-
Rutgers, Central Office	186362	186362	2000	common	421,537,000	1,189,937,000	35,682,000	-
Rutgers, Central Office	186362	186362	2001	common	437,081,000	1,261,644,000	38,052,000	-
Rutgers, Central Office	186362	186362	2002	gasb 34/35	436,602,000	1,248,159,000	57,063,000	1,687,307,000
Rutgers, Central Office	186362	186362	2003	gasb 34/35	436,122,000	1,314,121,000	59,517,000	1,795,135,000
Rutgers, Central Office	186362	186362	2004	gasb 34/35	435,347,000	1,410,163,000	68,214,000	1,879,574,000
Rutgers, New Brunswick	186380	186380	2005	gasb 34/35	455,033,000	1,465,054,000	76,682,000	1,946,335,000
Rutgers, New Brunswick	186380	186380	2006	gasb 34/35	495,934,000	1,548,195,000	77,401,000	2,040,826,000
Rutgers, New Brunswick	186380	186380	2007	gasb 34/35	455,764,000	1,659,656,000	79,843,000	2,219,394,000
Rutgers, New Brunswick	186380	186380	2008	gasb 34/35	476,511,000	1,687,708,000	80,959,000	2,310,436,000
Rutgers, New Brunswick	186380	186380	2009	gasb 34/35	450,134,000	1,767,729,000	95,414,000	2,309,115,000
Rutgers, New Brunswick	186380	186380	2010	gasb 34/35	444,009,000	1,870,938,000	107,168,000	2,426,600,000
Rutgers, New Brunswick	186380	186380	2011	gasb 34/35	438,156,000	1,985,979,000	115,595,000	2,636,769,000

PARID_F: Identification number for Parent Institution for Finance component
Account: Accountant standard used

How Does Parent-Child Reporting Affect Panel Construction and Analyses?

Parent-child reporting affects the construction and analyses of IPEDS panel data. Parent-child reporting creates difficulties for appending data over time within a survey component. Consider the goal of creating annual measures of state appropriations for each Rutgers University campus from the Finance data in Table 11.8. Researchers cannot know the state appropriations for each of the three campuses because finance variables were never reported separately for each campus. Furthermore, researchers cannot create annual organization-level measures by simply appending 1 year of data to the next because the parent institution changes over time.

Parent-child reporting makes it difficult to merge Fall Enrollment data in Table 11.7 to Finance data in Table 11.8 for several reasons. First, the child-level observations in the Fall Enrollment component contain UNITID-year observations that do not exist in the Finance component. Second, the observations for Rutgers Central Office (UNITID = 186362) will not merge with the Fall Enrollment component because the Fall Enrollment component does not contain data for Rutgers Central Office. Third, the New Brunswick campus observations (UNITID = 186380) from the Finance component will merge with observations from the Fall Enrollment component; however, this merge will be incorrect because the New Brunswick observations from the Finance component are parent-level observations, representing data from multiple campuses, but the New Brunswick observations from the Fall Enrollments component are child-level observations, representing data from one campus.

Parent-child relationships often affect whether panel analyses can be conducted using the desired unit of analysis (e.g., system level, organization level, or campus level). Panel analyses that utilize data from one survey component—other than Finance—can often use a campus-level unit of analysis for multicampus Title IV institutions that report separate data for each Title IV branch campus. For example, researchers can conduct campus-level analyses of change over time in nonresident enrollments at Rutgers University because Fall Enrollment data are reported at the campus level. However, panel analyses requiring variables from multiple survey components (especially Finance) often cannot use a campus-level unit of analysis. For example, analyses of the relationship between state appropriations and nonresident enrollments at Rutgers University cannot be conducted at the campus-level because Fall Enrollment data are reported at the campus level but Finance data are reported at the organization level.

In particular, parent-child relationships also affect the unit of analysis that can be utilized in panel analyses of for-profit institutions. The Title IV for-profit sector consists of a large number of Title IV institutions with one or a few campuses, and a smaller number of Title IV institutions with many campuses that operate across different states (e.g., University of Phoenix). Multicampus for-profit institutions often report campus-level data for some survey components (e.g., 12-Month

Enrollments) but organization-level data for other survey components, especially Finance. Therefore, panel analyses of for-profit institutions that require finance variables must often use an organization-level unit of analysis rather than a campus-level unit of analysis.

Patterns of Parent-Child Relationships

Understanding the patterns of parent-child relationships is a first step to overcoming problems associated with parent-child reporting. This section describes the extent of parent-child reporting, how parent-child relationships have changed over time, and which types of institutions are most likely to be effected by parent-child reporting. We focus mostly on the Finance component because parent-child reporting is the most prevalent and the most complicated in the Finance component.

As a general overview, parent-child reporting requirements changed in 2003–2004 (Association for Institutional Research 2003). Prior to 2003–2004, Title IV branch campuses often reported IC data but reported some or all of the remaining survey components at the parent level (U.S. Department of Education. NCES 2013d). Beginning in 2003–2004, multicampus Title IV institutions were asked to identify Title IV branch campuses that would not report any IPEDS data. These branch campuses appear in the NCES College Navigator website but do not appear in IPEDS data downloaded from the IPEDS Data Center (Association for Institutional Research 2013). However, for each Title IV branch campus that reported separate IPEDS data, it was required to complete all IPEDS survey components except Finance (Association for Institutional Research 2003).

Patterns of parent-child relationships can be described using parent-child “flag” variables. Flag variables exist for each survey component (e.g., Finance, Fall Enrollments) and identify the parent-child status (e.g., parent, child, neither) for that particular survey component. Table 11.9 presents the parent-child relationships over time for the Finance survey component.¹⁷

The first six columns of Table 11.9 indicate the parent-child Finance relationships from 1986–1987 to 2010–2011 using data derived from the Finance component. In 1986–1987, there were 3,604 observations in the Finance dataset (one observation per unique UNITID), 77 observations were parents, and 234 observations were children. Beginning in 1994–1995, the parent-child Finance flags indicated whether a child was a full child or a partial child. A full child refers to a child that reports all Finance data with the parent. Note that there were many full-child observations in data from the Finance component because full-child observations typically completed the IC component but not the Finance component. A partial child refers to a campus that reports

¹⁷ Non-degree-granting private nonprofit and private for-profit institutions (sectors 8 and 9) did not complete the Finance component for academic years 1986–1987 to 1998–1999. The number of observations in Finance data increased dramatically in 1999–2000 when a separate Finance survey form was created for for-profit institutions (Fuller 2011).

some variables at the child-level and some variables at the parent-level. The number of partial-child observations increased from 257 in 1999–2000 to 960 in 2010–2011.

The remaining five columns of Table 11.9 dataset provide parent-child Finance relationships from 1999–2000 to 2010–2011, an IC component dataset that has variables about parent-child relationships for the Finance survey component. Beginning with the year 2000 IPEDS collection, parent-child flags for all survey components were included in the Response Status data file associated with the IC component. Note that the number of partial-child observations using data from the IC component is about the same as the number of partial-child observations using data from the Finance component (first six columns of Table 11.9) because branch campuses completed both the IC component and the Finance component. The number of full-child observations using data from the IC component was 537 in 1999–2000 and 440 in 2010–2011, but no full-child observations exist over the time period using data from the Finance component because these branch campuses (e.g., Rutgers University, Newark) did not complete the Finance component.

The distinction between full-child and partial-child campuses is intimately related to Program Participation Agreements (PPA) and accounting standards. Accounting standards are discussed in greater detail in the Common Data Challenges section of this chapter. Briefly, all institutions used the “Common Form” Finance survey from 1974–1975 to 1995–1996. Beginning in 1996–1997, the Finance component was revised for private not-for-profit institutions due to changes in accounting standards released by the Financial Accounting Standards Board (FASB). Beginning in 1999–2000, a FASB accounting form was created for for-profit institutions. Beginning in 2001–2002, the Governmental Accounting Standards Board (GASB) Finance form was phased-in for public institutions in order to satisfy new accounting standards implemented by GASB.

Title IV branch campuses (e.g., Rutgers University–Newark) that share a PPA with a Title IV main campus (e.g., Rutgers University–New Brunswick) have the choice of reporting all Finance data at the child level, reporting some Finance data at the parent level (partial child), or reporting all Finance data at the parent level (full child) (U.S. Department of Education. NCES 2012).

A Title IV institution with a separate PPA is not allowed to be a full child; however, Title IV institutions that are part of a system sometimes have difficulty reporting asset, liability, and equity variables separately for each Title IV institution in the system (U.S. Department of Education. NCES 2012). The Common Form Finance survey did not contain many survey items about assets, liabilities, and equity (U.S. Department of Education. NCES 2000). However, FASB accounting standards and GASB accounting standards require more information about assets, liabilities, and equities, making it difficult for some Title IV institutions within a system to answer all Finance survey items at the Title IV institution level (U.S. Department of Education. NCES 2012).

Therefore, Title IV institutions that are part of a system are allowed to report some variables at the system level (i.e., be a partial child). In particular, public institutions using GASB standards are allowed to report statement of net assets (Part A), summary of changes in net assets (Part D), debt and assets (Part L), and endowment assets (Part H) at the system level but must report revenues (Part B), expenses (Part C), and scholarships

and fellowships (Part E) separately at each Title IV institution (U.S. Department of Education. NCES 2012). Private nonprofit institutions using FASB standards may report statements of financial position (Part A), changes and net assets (Part B), and endowment assets (Part H) at the system level but must report scholarships and fellowships (Part C), revenues (Part D), and expenses (Part E) separately for each Title IV institution (U.S. Department of Education. NCES 2012). For-profit institutions may report balance sheet (Part A) and change in equity (Part B) at the system level but must report student grants (Part C), revenues (Part D), and expenses (Part E) separately for each Title IV institution (U.S. Department of Education. NCES 2012).

Relative to the Finance component, parent-child relationships for other survey components are simpler because children are always full children; partial children do not exist. Table 11.10 displays parent-child flags for the Completions and Fall Enrollments components, respectively. The first four columns of Table 11.10 display parent-child Completions flags using data from the Completions survey component, and the final four columns indicate the parent-child Completions flags using data from the IC survey component. Starting in 2003–2004 IPEDS required all Title IV branch campuses that report IPEDS data to complete all survey components except Finance (U.S. Department of Education. NCES 2013d). Examining the parent-child Completions flags from the IC component, the number of parents decreased from 83 in 2002–2003 to 43 in 2003–2004 to 9 in 2010–2011, and the number of children decreased from 165 in 2002–2003 to 64 in 2003–2004 to 13 in 2010–2011. Table 11.10 indicates that there are similar parent-child patterns for the Fall Enrollment component. In contrast, the number of parent-child relationships in the Finance component is much greater than the number of parent-child relationships in other survey components.

What types of organizations have been affected most often by these complexities in parent-child relationships? Table 11.11 displays parent-child Finance flags by selected sector for the 2000–2001, 2005–2006, and 2010–2011 academic years using data from the IC component. Table 11.11 excludes campuses ineligible for Title IV financial aid, using a Title IV flag indicator available beginning with IC data collected in fall 1997 (which refers to Finance data for 1995–1996). Table 11.11 also excludes administrative offices, which have sector=0.

For public four-year institutions (sector 1; sample size=649 in 2000–2001), the number of full-child observations increased from 17 in 2000–2001 to 36 in 2010–2011, whereas the number of partial-child observations decreased from 85 in 2000–2001 to 39 in 2010–2011. For private nonprofit four-year institutions (sector 2; sample size=1,620 in 2000–2001), the number of full-child observations decreased from 94 in 2000–2001 to 50 in 2010–2011, whereas the number of partial-child observations increased from 27 in 2000–2001 to 53 in 2010–2011. For public two-year institutions (sector 4; sample size=1,180 in 2000–2001), the number of full-child observations decreased from 109 in 2000–2001 to 23 in 2010–2011, whereas the number of partial-child observations increased from 54 in 2000–2001 to 148 in 2010–2011. Finally, at for-profit institutions (sectors 3, 6, and 9; sample size=2,503 in 2000–2001 and 3,454 in 2010–2011), the number of full-child observations decreased from 337 in 2000–2001 to 299 in 2010–2011, whereas the number of partial-child observations increased from 170 in 2000–2001 to 469 in 2010–2011.

Table 11.10 Parent-child indicator for Completions component and Fall Enrollments component, academic years 1986–1987 to 2010–2011

	Parent-child flags for Completions component				Parent-child flags for Fall Enrollments component			
	From Completions data		From IC data		From Fall Enrollments data		From IC data	
	NA/miss	Parent	Child	Parent	NA/miss	Parent	Child	Parent
1987	4,874	51	0		6,485	52	0	
1988	5,564	42	1		6,449	59	0	
1989	5,757	32	0		5,324	31	0	
1990	4,852	25	0		5,564	26	0	
1991	4,995	19	0		5,273	23	0	
1992	5,015	27	0		6,464	15	0	
1993	5,109	48	0		6,589	28	0	
1994	5,091	63	0		8,775	52	0	
1995	8,442	103	167		8,494	88	0	
1996	6,554	104	0		8,378	107	0	
1997	6,373	113	0		6,566	100	0	
1998	7,087	86	0		6,359	102	0	
1999	6,151	76	0		7,053	84	0	
2000	6,571	99	0	100	6,144	72	0	
2001	6,435	101	0	101	6,476	80	0	9,347
2002	6,386	87	0	87	6,469	91	0	80
2003	6,498	83	0	83	6,457	69	0	91
2004	6,594	41	0	41	6,508	67	1	68
2005	6,696	21	0	21	6,612	17	0	67
2006	6,748	19	0	19	6,679	10	0	17
2007	6,769	16	0	16	6,745	3	0	10
2008	6,898	13	0	13	6,757	4	0	3
2009	7,018	11	0	11	6,885	1	0	4
2010	7,259	10	0	10	6,984	6	0	1
2011	7,480	9	0	9	7,241	5	0	12
								6
								5
								7,382
								6
								137
								179
								144
								138
								25
								9
								3
								4
								1
								6
								5

Table 11.11 Parent-child Finance flags 2000–01, 2005–06, and 2010–11 by selected sector using IC data, excluding non-Title IV campuses

	Public, 4-year or above (sector=1)					Private nonprofit, 4-year or above (sector=2)				
	Neither	Parent	Full- child	Partial- child	Total	Neither	Parent	Full- child	Partial- child	Total
2001	524	23	17	85	649	1,449	50	94	27	1,620
2006	573	23	33	32	661	1,459	30	63	48	1,600
2011	601	20	36	39	696	1,481	26	50	53	1,610
	Public, 2-year (sector=4)					Private, for profit, (sectors 3, 6, 9)				
	Neither	Parent	Full- child	Partial- child	Total	Neither	Parent	Full- child	Partial- child	Total
2001	991	26	109	54	1,180	1,786	204	337	176	2,503
2006	944	11	75	131	1,161	2,039	170	189	330	2,728
2011	906	5	23	148	1,082	2,510	176	299	469	3,454

Collapsing Solutions to Parent-Child Reporting

Example of a Collapsing Solution

Parent-child reporting makes it difficult to create an IPEDS panel dataset with one observation per organization-year (or one observation per campus-year) that has variables from multiple IPEDS survey components. It is important to devise solutions to parent-child reporting that are relevant to the research question at hand. We discuss two general solutions, collapsing and allocation. The former method collapses child observations into parent observations separately for each survey component. The collapsed panel datasets for each survey component can then be merged to create an analysis dataset. The allocation solution allocates variables reported at the parent level to the child level so that the number of observations does not decrease.

We demonstrate a simple collapsing solution using Fall Enrollment and Finance data from the North Orange County (California) Community College District, which is comprised of Fullerton College (UNITID=114859), Cypress College (UNITID=113236), and the North Orange County Community College District Office (UNITID=120023). Table 11.12 provides Fall Enrollment variables (full-time student headcount and part-time student headcount) from 2001–2002 to 2006–2007. Fall Enrollment data are reported separately at Fullerton College and Cypress College, but no observations exist for the District Office. Table 11.13 provides selected finance variables from 2001–2002 to 2006–2007. From 2001–2002 to 2003–2004, all finance variables are reported at the parent level (the District Office), and there are no observations for Cypress College or Fullerton College. In 2004–2005 and 2005–2006, all finance variables are reported at the campus level, and there are no observations for the District Office. In 2005–2006 and 2006–2007, revenue variables (state appropriations and total current revenues) are reported at

Table 11.12 Fall enrollment data for North Orange County Community College District, prior to collapse

Institution	PARID	UNITID	Year	Undergrad full-time	Undergrad part-time
Fullerton College	114859	114859	2002	7,236	13,692
Cypress College	114859	113236	2002	4,951	8,945
Fullerton College	114859	114859	2003	7,817	13,106
Cypress College	114859	113236	2003	5,019	10,909
Fullerton College	114859	114859	2004	7,605	11,115
Cypress College	114859	113236	2004	4,881	7,800
Fullerton College	114859	114859	2005	8,079	11,695
Cypress College	114859	113236	2005	5,193	8,106
Fullerton College	114859	114859	2006	8,101	11,510
Cypress College	114859	113236	2006	4,933	17,032
Fullerton College	114859	114859	2007	8,245	11,750
Cypress College	114859	113236	2007	4,863	8,035

Undergrad full-time: total of undergrad students full-time

Undergrad part-time: total of undergrad students part-time

PARID: Identification Number of Parent Institution

the campus level, but asset variables (land value and net assets) are reported at the district level.

The first step in a collapsing solution is to choose the parent. We chose Fullerton College to be the parent because Fullerton College had higher enrollments than Cypress College in most years, but either choice is reasonable because neither campus is obviously a main campus. Choosing an administrative unit (e.g., North Orange County Community College District Office) to be the parent has strengths and drawbacks, as we discuss below.

The second step is to “collapse” variables measured at the child level to the parent level. This step must be conducted separately for each survey component. “Collapsing” means creating “output” parent-year observations that are the sum of all parent- and child-year “input” observations. Table 11.14 provides collapsed Fall Enrollment data for 2002–2007, which has one observation per year. The institutional identification code refers to Fullerton College, which we chose as the parent institution, but Table 11.14 contains Fall Enrollment data from both Fullerton College and Cypress College. For example, the value of undergraduate full-time enrollments in 2005 was 13,272 in Table 11.14, which is the sum of undergraduate full-time enrollments from Fullerton College (8,079) and Cypress College (5,193) in Table 11.12.

Table 11.15 provides collapsed Finance data from 2002 to 2007. The institution identification code refers to Fullerton College, but Table 11.15 contains Finance data from Fullerton College, Cypress College, and the District Office. For example, the value of 2006 state appropriations is \$73,636,315 in Table 11.15, which is the sum of 2006 state appropriations from Fullerton College (\$46,390,878), Cypress College (\$27,245,437), and the District Office (missing) from Table 11.10. The

Table 11.13 Finance data for North Orange County Community College District, prior to collapse

Institution	PARID	UNITID	Year	State approp.	Current rev	Land	Net assets
North Orange County Cc District	114859	120023	2002	42,270,926	178,950,478	13,566,862	130,731,217
North Orange County Cc District	114859	120023	2003	37,198,015	183,820,724	13,955,737	145,015,093
North Orange County Cc District	114859	120023	2004	43,209,813	180,175,915	14,943,597	182,429,509
Fullerton College	114859	114859	2005	37,903,411	116,583,456	9,055,820	125,621,107
Cypress College	114859	113236	2005	24,643,472	75,798,483	5,887,777	81,674,449
Fullerton College	114859	114859	2006	46,390,878	128,774,366	—	—
Cypress College	114859	113236	2006	27,245,437	75,629,390	—	—
North Orange County Cc District	114859	120023	2006	—	—	14,943,597	233,682,281
Fullerton College	114859	114859	2007	53,860,373	153,860,154	—	—
Cypress College	114859	113236	2007	33,011,196	94,301,384	—	—
North Orange County Cc District	114859	120023	2007	—	—	17,453,683	295,342,213

PARID: Identification Number of Parent Institution

Table 11.14 Fall enrollment data for North Orange County Community College District, after collapse

Institutions	UNITID	Year	Undergrad full-time	Undergrad part-time
Fullerton College	114859	2002	12,187	22,637
Fullerton College	114859	2003	12,836	24,015
Fullerton College	114859	2004	12,486	18,915
Fullerton College	114859	2005	13,272	19,801
Fullerton College	114859	2006	13,034	28,542
Fullerton College	114859	2007	13,108	19,785

Undergrad full-time: total of undergrad students full-time

Undergrad part-time: total of undergrad students part-time

Table 11.15 Finance data for North Orange County Community College District, after collapse

Institution	UNITID	Year	State approp.	Current rev.	Land	Net assets
Fullerton College	114859	2002	42,270,926	178,950,478	13,566,862	130,731,217
Fullerton College	114859	2003	37,198,015	183,820,724	13,955,737	145,015,093
Fullerton College	114859	2004	43,209,813	180,175,915	14,943,597	182,429,509
Fullerton College	114859	2005	62,546,883	192,381,939	14,943,597	207,295,556
Fullerton College	114859	2006	73,636,315	204,403,756	14,943,597	233,682,281
Fullerton College	114859	2007	86,871,569	248,161,538	17,453,683	295,342,213

value of 2006 net assets is \$233,682,281 in Table 11.15, which is the sum of 2006 state appropriations from Fullerton College (missing), Cypress College (missing), and the District Office (\$233,682,281) in Table 11.13. The final step of the collapse solution is to merge the collapsed Fall Enrolment data and the collapsed Finance data, as demonstrated in Table 11.17.

Having provided a simple example of collapsing an institution for one parent-child relationship (i.e., North Orange County Community College District), we next summarize the main steps in collapsing *all* child-level observations into parent-level observations. The first step is to create separate panel datasets of non-collapsed data for each survey component (as demonstrated in Tables 11.12 and 11.13 in the North Orange County example). The second step is to create a parent-child list that identifies the UNITID of the parent associated with the UNITID for each child. In the North Orange County example, we chose Fullerton College (UNITID=114859) to be the parent campus and Cypress College (UNITID=113236) and the District Office (UNITID=120023) to be the child campuses. The parent-child list would specify the parent campus and the child campuses for all parent-child relationships. The third step is to merge the parent-child list to the panel

dataset for each survey component. The fourth step is to collapse the panel datasets for each survey component to the parent level, as shown in Tables 11.14 and 11.15. The fifth step is to merge the collapsed panel datasets from each survey component by UNITID (i.e., the UNITID of the parent) and year, as shown in Table 11.16.

The fourth and fifth steps are somewhat different for the IC survey component. Panel datasets of IC data should not be collapsed to the parent level because most IC variables (e.g., highest award level, zip code) cannot be summed across institutions. When merging IC data to collapsed data from other survey components, researchers should drop “child” observations from IC that do not merge with the parent-level observations from other survey components. However, researchers should be aware that merging campus-level observations from IC (i.e., not collapsed to the parent level prior to merging) to parent-level observations from other survey components means that the IC data attached to each parent observation in the analysis dataset is not representative of campuses in the collapsed record. We will demonstrate that this issue has important effects for research using the DCP Database.

Parent-Child Lists

The most important and most time-consuming step in implementing a collapsing solution is creating the parent-child list. A parent-child list is essentially a dataset with two important variables, “parent_UNITID” and “child_UNITID.” Table 11.17 displays selected observations from one potential parent-child list. Note that each child_UNITID is unique and that the UNITID of the parent is also included as an observation in the child_UNITID variable.

Although one collapsing solution may differ from another, all collapsing solutions follow two general rules that relate to the parent-child list. First, for a given parent-child relationship, the designation of which campus is the parent and which campuses are children must be constant over time and across survey components. This rule is enforced by the fact that each child_UNITID in the parent-child list is unique and is associated with only one parent_UNITID. In the example of Finance data at the Rutgers University campus (Table 11.8), the New Brunswick campus was the parent in some years and the Central Office was the parent in the other years, but the parent-child list allows only one Rutgers University campus to be the parent.

The second general rule of collapsing is that if a child campus *ever* reports an analysis variable of interest at the parent level for a particular survey component, then the child campus must be collapsed to the parent level for all years, and all survey components, and all variables. This rule is enforced by the fact that (1) the same parent-child list is merged to the panel dataset for each survey component (i.e., parent-child list is the same across survey components) and (2) the parent-child list is merged to the panel datasets for each survey component by child_UNITID rather than by child_UNITID and year (i.e., parent-child relationships are constant over time). For example, state appropriations for North Orange County campuses were not reported at the child level for some years (Table 11.13);

Table 11.17 Selected observations from a master parent-child list

Parent institution	Child institution	UNITID_parent	UNITID_child	OPEID_child
The University Of Alabama	The University Of Alabama	100751	100751	001051-00
The University Of Alabama	University Of Alabama Gadsden Center	100751	102401	
Faulkner University	Faulkner University	101189	101189	001003-00
Faulkner University	Faulkner University-Birmingham	101189	101198	001003-01
Faulkner University	Faulkner University-Florence	101189	101204	001003-02
Faulkner University	Faulkner University	101189	101213	001003-03
Faulkner University	Faulkner University-Mobile	101189	261418	001003-04
Glendale Community College	Glendale Community College	104708	104708	001076-00
Glendale Community College	Maricopa Community College System Office	104708	105136	001075-00
Glendale Community College	Gateway Community College	104708	105145	008303-00
Glendale Community College	Mesa Community College	104708	105154	001077-00
Glendale Community College	Phoenix College	104708	105428	001078-00
Glendale Community College	Rio Salado College	104708	105668	021775-00
Glendale Community College	Scottsdale Community College	104708	105747	008304-00
Glendale Community College	South Mountain Community College	104708	105792	021466-00
Glendale Community College	Paradise Valley Community College	104708	364016	026236-00
Glendale Community College	Chandler/Gilbert Community College	104708	364025	030722-00
Glendale Community College	Estrella Mountain Community College	104708	384333	031563-00
City College Of San Francisco	City College Of San Francisco	112190	112190	004502-00
City College Of San Francisco	San Francisco Community College Skills Center	112190	122481	
City College Of San Francisco	San Francisco Community College District	112190	122621	
Jones College-Jacksonville	Jones College-Jacksonville	135063	135063	001497-00
Jones College-Jacksonville	Jones College-Miami Campus	135063	402633	001497-10

Table 11.18 Sample size for selected years by 2005 Carnegie classification and control

	DCP Database									
	Public institutions					Private non-profit institutions				
	Res/Doct	Mast	Bach	Assoc	Specialized	Res/Doct	Mast	Bach	Assoc	Specialized
1987	152	241	63	822	65	90	319	493	126	534
1990	152	241	64	828	68	90	319	493	128	542
1995	152	243	65	837	72	90	319	495	134	570
2000	152	244	65	837	72	90	317	489	128	562
2005	152	244	66	835	68	88	313	481	113	532
2010	152	242	66	833	64	88	311	474	103	505

	Jaquette & Parra Panel									
	Public institutions					Private non-profit institutions				
	Res/Doct	Mast	Bach	Assoc	Specialized	Res/Doct	Mast	Bach	Assoc	Specialized
1987	165	267	79	908	69	90	321	485	114	505
1990	165	267	80	910	70	91	319	484	119	519
1995	165	268	79	922	77	91	318	488	127	538
2000	165	269	80	924	79	90	318	487	136	550
2005	165	270	81	922	71	88	313	481	115	515
2010	165	269	81	917	67	88	310	470	103	496

therefore, we collapsed child-level observations (Fullerton College, Cypress College, and the District Office) into parent level (Fullerton College) for all years. We conducted this step separately for both the Finance component and the Fall Enrollments survey component, even though Fall Enrollment data were reported at the child level for all years.

How does a researcher actually create a parent-child list? For each survey component and year, parent-child flag variables identify the presence of a parent-child relationship and parent-id variables identify the UNITID of the parent associated with each child. The first step in creating a parent-child list is to identify and group together all campuses that have *ever* been in a parent-child relationship during the desired analysis period in a survey component that is used to create analysis variables. This step is completed using parent-child flag variables and parent-id variables.¹⁸ Researchers should be aware that parent-child flag variables and parent-id variables from a single year of data (e.g., 2002–2003) cannot be used to create a parent-child list for an analysis dataset that spans multiple years of data.

For example, if a researcher is creating an analysis dataset for the academic years 2000–2001 to 2010–2012 using variables from IC, Finance, and 12-Month

¹⁸ This step should be conducted separately for each survey component. With respect to the Finance component, the researcher would sort un-collapsed Finance data by UNITID and parent_UNITID (idx_f) and “fill down” the parent_UNITID variable for all years of data for campuses that have ever had a non-missing parent_UNITID variable. The variable parent_UNITID should be equal to UNITID for any campus that has ever been a parent. Finally, the researcher can group all observations that have ever been in the same parent-child relationship (for a particular survey component) by sorting by parent_UNITID, year, and UNITID.

Enrollments, then the parent-child list should be created using Finance data and 12-Month Enrollment data from 2000–2001 to 2010–2012. Additionally, if the analysis dataset uses revenue and expenditure variables but not asset and liability variables, then the parent-child list should not include partial children that always reported revenue and expenditure variables at the child level during the analysis period.

Once all observations that have ever been in a parent-child relationship are identified and grouped together, the researcher must decide which UNITID will be the parent and which UNITIDs will be the child campuses. When the list of observations in a parent-child relationship does not include an administrative unit (e.g., Maricopa Community College System Office), researchers can usually identify a main campus to choose as a parent. Some decisions are obvious, as in the case of University of Alabama in Table 11.17. Other decisions require investigation, as in the case of Jones College in Table 11.17.

Selecting the parent is less obvious when the list of observations in a parent-child relationship includes an administrative unit. If the researcher wants to avoid selecting the administrative office as the parent, it may be that none of the remaining campuses are main campuses (e.g., California Community College districts) and the researcher should just choose one campus to be the parent. Choosing administrative units (e.g., North Orange County District Office, Rutgers University Central Office) as parents has both drawbacks and strengths. The drawbacks are that administrative units have many missing IC variables (e.g., 25th percentile SAT score, tuition, and fees) and cannot be easily merged to other data sources (e.g., Carnegie Classification). The strength is that identifying the collapsed observations as an administrative unit consisting of multiple campuses is more accurate than assigning the collapsed observations the IC data associated with one particular campus.

The sample parent-child list in Table 11.17 also displays the OPEID for each child. An alternative approach is to create a parent-child list based on the first six digits of OPEID. The rationale for this approach is that each Title IV institution with a separate Program Participation Agreement (PPA) must complete all IPEDS survey components but Title IV branch campuses do not have to complete all IPEDS survey components (U.S. Department of Education. NCES 2012). In theory, the first six digits of OPEID could be used to identify all campuses in a parent-child relationship, and the two-digit suffix of OPEID could be used to assign the parent (suffix=00) and the child observations (suffix≠00). A benefit of this approach is that using six-digit OPEID to identify parent institutions would make it easier to merge IPEDS data to data from the Office of Postsecondary Education and Federal Student Aid, which we discuss below. We have not yet attempted this approach because we only recently learned about the relationship between PPA, OPEID, and parent-child reporting.

However, using OPEID to create parent-child lists involves at least two difficulties. First, the IC component began collecting OPEID in 1995–1996 and some children—for example, the University of Alabama–Gadsden Center in Table 11.17—never received an OPEID because they last reported IPEDS data prior to 1995–1996. This problem makes it difficult to identify child campuses using OPEID alone. Second,

parent-child relationships can involve multiple Title IV institutions when some finance variables (e.g., assets) are reported at the system level (U.S. Department of Education, NCES 2012). In Table 11.17, each community college in the Maricopa Community College System is a Title IV institution with a separate PPA, but some Finance data has been reported at the system level. The OPEID variable is not effective for identifying parent-child relationships that span Title IV institutions. Although both of these difficulties are tractable, researchers should be aware that creating a parent-child list based on OPEID may require triangulating several additional sources of data (e.g., parent-child flags, parent-UNITID variables).

In the following section, we analyze the collapsing solution used by the Delta Cost Project. As mentioned before, the DCP Database is a premade panel dataset that incorporates IPEDS data containing variables from Finance, Fall Enrollments, 12-Month Enrollments, Human Resources, Graduation Rates, and Student Financial components. Because of its popularity, it is important to understand the rationale behind the DCP's parent-child collapsing solution and the potential problems that can ensue when applying DCP to study particular research questions.

The Delta Cost Project Collapsing Solution

The DCP Database used a collapsing solution to create a panel dataset using data from multiple IPEDS survey components. DCP used the following rule to decide whether to collapse campuses: "Institutions that reported data together (on any of the IPEDS surveys) for any year starting in 1987 have been grouped together for all years to maintain consistency of the data over time" (Delta Cost Project n.d.).¹⁹ The DCP parent-child list shown in Delta Cost Project (n.d.) has 222 parent institutions linked to 593 child institutions. These 815 institutions (593 + 222) were collapsed into 222 observations per year. About 71 % of these 815 institutions are public, reflecting the tendency for public institutions to have multicampus institutions with centralized data reporting.

For many state systems, the DCP solution collapsed multiple Title IV institutions into a single observation representing the entire state system. For example, all University of Texas (UT) campuses (e.g., Austin, El Paso, San Antonio) were collapsed into a single observation. Other examples include all University of Alaska campuses, all University of Illinois campuses, all University of Maine campuses, all University of Massachusetts campuses, all University of Tennessee campuses, all University of Missouri campuses, and all Indiana Community College campuses. For each state system collapsed into a single observation, the DCP parent-child list typically defined the parent as the "flagship" campus (e.g., University of Texas–Austin and Pennsylvania State University–University Park). Therefore, UNITIDs that represent state systems on the DCP Database often have IC data associated with

¹⁹ At the time this manuscript was prepared, Delta Cost Project (n.d.) could not be downloaded from the DCP website.

the flagship campus, but Fall Enrollment, Finance, Completions, and other component data associated with the entire state system.

Additionally, DCP collapsed several substate systems of Title IV institutions into single observations. For example, each City University of New York (CUNY) campus (e.g., CUNY Graduate School and University Center, LaGuardia Community College) is a Title IV institution with a separate PPA. However, the DCP Database solution collapsed all 20 CUNY campuses into a single observation with the parent defined as CUNY City College.

The decision by DCP to collapse systems of Title IV institutions to the system level affects the kinds of research questions that can be answered using the DCP Database. The DCP Database should not be used to answer research questions that compare one type of public university to another, for example “what is the effect of banning affirmative action on underrepresented minority enrollments at public flagship campuses versus public regional campuses?” (e.g., research vs. master’s). In addition, the effect of state appropriations on nonresident enrollments cannot be reliably answered using the DCP Database because the observations for “flagship” campuses will include many “regional” campuses that have weak demand from nonresident students. Also, the DCP Database cannot be used to address research questions that compare public and private universities of the same type. For example, it cannot be used to compare expenditure per FTE at public research universities versus private research universities because the expenditure per FTE measures at public research universities will include data from non-research universities.

Collapsing Title IV institutions to the system level is unavoidable when desired finance variables are reported only at the system level. However, our preliminary investigation of the DCP Database suggests that some state systems did not need to be collapsed to the system level. For example, each University of Missouri campus reported revenue, expenditure, and asset variables in all years of data. The University of Missouri system office also reported finance variables (e.g., revenues, assets), but the amounts were small relative to those of the flagship Columbia campus. Nevertheless, the DCP collapsing solution collapsed all University of Missouri campuses to the system level.

Furthermore, the DCP Database collapsed institutions that were purchased by another institution in ways that may cause problems for some research applications. For example, Barat College was an undergraduate college for women founded in 1858 in Chicago. Facing severe financial problems, Barat College was purchased by DePaul College in 2000–2001. Barat College continued to report IPEDS data until 2004–2005 when the Barat College campus was sold by DePaul College to become a condominium development.²⁰ However, the DCP parent-child list identified Barat College as a child of DePaul University; therefore, the DCP Database collapsed data from Barat College into DePaul College for the years 1987–2001 even though the two institutions had no relationship during this time period. The DCP collapse solution treats other mergers similarly. It is inappropriate to collapse a purchased child institution into the purchasing parent institution prior to the point of sale.

²⁰ http://en.wikipedia.org/wiki/Barat_College

Collapsing Solutions for Specific Research Questions

The DCP parent-child list is not “wrong,” but it is not appropriate for all research questions because no single parent-child list is appropriate for all research questions. Rather than rely on a readily available panel dataset with a one-size-fits-all collapsing solution, researchers should develop collapsing solutions that are appropriate for the particular research question at hand.

When considering what—if any—collapsing solution to implement, researchers should consider what unit of analysis (e.g., campus level, organization level, system level) is appropriate, which IPEDS survey components will be utilized, and whether it is necessary to merge across IPEDS survey components. An organization-level unit of analysis (i.e., each observation representing a Title IV institution) requires researchers to implement some kind of collapsing solution because Title IV branch campuses often report IPEDS data. However, researchers selecting a campus-level unit of analysis should not collapse Title IV branch campuses to the Title IV institution level. A campus-level unit of analysis may be possible when analyses do not require merging data from multiple survey components.

In general, research questions that require an analysis dataset with data from multiple survey components (especially Finance) necessitate collapsing to the organization level or even the system level. Analyses that require asset and liability measures may require collapsing to the system level for Title IV institutions that reported the desired measures at the system level. However, analyses that utilize revenue/expenditure/scholarship measures, but not asset/liability measures, typically do not have to collapse Title IV institutions to the system level because revenue/expenditure/scholarship measures are generally reported separately for each Title IV institution. The DCP Database collapsed many Title IV institutions to the system level to create asset/liability variables even though researchers often use the DCP Database to conduct analyses that do not require asset/liability measures.

To illustrate the idea that collapsing solutions should be appropriate for the research question at hand, we answer a simple research question using the DCP Database and an alternative IPEDS panel we constructed. The research question is, “How does instructional expenditure per student vary across different types of public institutions?”

The motivation for this research question is that public universities have responded to declining state appropriations by attempting to generate “alternative” revenue sources (e.g., research funding, donations, nonresident students). However, research suggests that only flagship public universities can generate substantial revenues from these sources (Cheslock and Gianneschi 2008; Jaquette and Curs 2013; Slaughter and Leslie 1997; Winston 2003). Widening differentials in spending per student at flagship universities relative to regional universities and community colleges may undermine the policy goal of equality of opportunity.

The Jaquette-Parra Panel was created using a collapsing solution that is adequate to answer the above research question. Because answering this research question requires only expenditure measures from the Finance component, we did

not collapse partial-child observations to the parent level so long as all the campuses within the parent-child relationship always reported expenditure data at the child level. Therefore, we did not collapse Title IV institutions to the system level when they reported asset/liability variables at the system level but also reported expenditure variables at the Title IV institution level. Additionally, we did not collapse child campuses that failed to report expenditure data to the parent level if we perceived those campuses to represent an entire state system. For example, we did not collapse Louisiana Technical College campuses or Pennsylvania State University campuses to the parent level. Had we understood the difference between a multicampus Title IV institution (e.g., Pennsylvania State University) and a state system (e.g., University of Texas System) more completely at the time we created the collapsing solution, we would have been willing to collapse any multicampus Title IV institutions.

Table 11.18 shows the sample sizes of public institutions by 2005 Carnegie Classification for the DCP Database and the Jaquette-Parra Panel. Compared to the Jaquette-Parra Panel, the DCP Database has fewer observations for public doctoral/research (Res/Doct), public master's (Mast), and public baccalaureate (Bach) because many distinct Title IV institutions are collapsed to the state system level and were assigned the parent-UNITID associated with the flagship campus. The DCP Database also has fewer public associate's (Assoc) observations because community colleges in several state systems (e.g., Kentucky, Louisiana, and Indiana) were collapsed into a single observation. However, the DCP decision to collapse all Louisiana and Kentucky community colleges to the state level *is* warranted because most of the campuses in these systems are Title IV branch campuses rather than Title IV institutions.

Table 11.19 compares "education and related expenditures per FTE student" in the DCP Database to the Jaquette-Parra Panel measures for the same across selected years. This crude measure of E&R per FTE was calculated by adding total expenditures on instruction, student services, and academic support and dividing it by a measure of FTE for students from the Fall Enrollment component (that counted a part-time student as 1/3 of a full-time student). Values of E&R per FTE in the DCP Database and the Jaquette-Parra Panel are identical for institutions that were not collapsed. However, compared to the Jaquette-Parra Panel, mean E&R per FTE at Res/Doct from the DCP Database is much lower because many flagship public universities in the DCP Database contain data from both flagship and regional universities. Similarly, the ratios of mean Res/Doct expenditure to mean expenditure at Mast, Bach, and Assoc institutions are much lower in the DCP Database than the Jaquette-Parra Panel.

Policy researchers using the DCP Database to investigate differentials in spending per student between public flagship and public regional universities may, therefore, underestimate the spending differential between public research universities and public master's/bachelor's institutions because observations for research universities include data from master's and bachelor's institutions. However, the Jaquette-Parra collapsing solution is also imperfect (e.g., we should have collapsed Pennsylvania State University to the Title IV institution level).

Table 11.19 Mean education and related expenditure (E&R) per FTE at public institutions, by 2005 Carnegie Classification

	DCP Database				Ratio of mean Res/ Doct to mean:		
	Mean E&R per FTE				Doct to mean:		
	Res/Doct	Mast	Bach	Assoc	Mast	Bach	Assoc
2004	13,152	8,703	8,060	6,951	1.51	1.63	1.89
2005	13,409	8,761	7,930	7,000	1.53	1.69	1.92
2006	13,643	8,871	8,385	7,200	1.54	1.63	1.89
2007	13,946	9,057	8,572	7,477	1.54	1.63	1.87
<i>N</i> (2005)	152	244	66	835			
<i>N</i> non-missing (2005)	151	242	63	824			
	Jaquette-Parra Panel				Ratio of mean Res/ Doct to mean:		
	Mean E&R per FTE				Doct to mean:		
	Res/Doct	Mast	Bach	Assoc	Mast	Bach	Assoc
2004	14,186	8,739	8,336	7,101	1.62	1.70	2.00
2005	14,470	8,786	8,348	7,161	1.65	1.73	2.02
2006	14,712	8,894	8,624	7,331	1.65	1.71	2.01
2007	15,063	9,071	8,923	7,616	1.66	1.69	1.98
<i>N</i> (2005)	165	270	81	922			
<i>N</i> non-missing (2005)	164	265	81	884			

Allocation Solutions to Parent-Child Reporting

Rather than collapsing child institutions into parent institutions, allocation solutions allocate data reported at the parent level to child institutions. Allocation solutions appear particularly useful when variables for a state system are reported at the parent level such that collapsing solutions would collapse an entire state system into a single observation. For example, the University of Alaska System has three campuses (Fairbanks, Anchorage, and Southeast). Fall Enrollment data and Completions data are reported separately for each campus in all years. Revenue and expenditure data are also reported at the campus level, but many asset variables (e.g., market value of endowment) are reported at the system office level. The allocation solution leaves measures of Fall Enrollment, Completions, and revenues and expenditures untouched, but asset variables reported at the parent level are allocated to the child level based on some predetermined rule. Whereas the collapsing solution results in one observation per year, the allocation solution results in three observations per year (ignoring the system office). If a reasonable allocation rule is available, then an allocation solution is preferable to a collapsing solution because distinct campuses (e.g., University of Alaska Fairbanks, University of Alaska Anchorage) remain distinct observations in the analyses.

The most critical step in an allocation solution is choosing an “allocation factor variable” that identifies what proportion of a parent-level variable to allocate to each campus. For example, in 2010 the University of Alaska System Office reported an end-of-year market value of their endowment of \$182 million. What proportion of this \$182 million should be allocated to the Fairbanks, Southeast, and Anchorage campuses? IPEDS reports allocation factor flags for each survey component that identify the proportion of parent-level variables that should be allocated to each child campus. However, we have found the IPEDS allocation factor flags unhelpful because they are unavailable prior to 2004 and each organization uses a different metric to allocate parent-level data to child campuses; often, the allocation flags allocate 100 % of parent-level data to a “main campus.” Alternatively, a researcher could allocate parent-level variables based on the proportion of total enrollments, current revenues, etc. at each campus.

Although the allocation solution is attractive in theory, we have found it to be infeasible in practice. Relative to collapsing solutions, the programming for allocation solutions is arduous. Collapsing solutions only require the creation of a parent-child list, and then all variables are easily collapsed to the parent level in all years and in all survey components. By contrast, allocation solutions require allocating specific variables from the parent level to the child level, but the designation of the parent may change from year to year (e.g., the main campus or the system office) and in some years the variable may be reported at the parent level rather than the child level. Therefore, allocation solutions must be done one parent-child relationship at a time and one variable at a time. We spent hundreds of hours creating a programming solution that could allocate parent-level variables to child campuses on the basis of allocation factor variables of our choosing (e.g., percent of total enrollments, percent of total revenues, percent of total assets), which could change from year to year.

Although the programming for allocation solutions is tedious, the primary reason for rejecting allocation solutions is the inherent arbitrariness of any allocation factor chosen. For example, imagine that you are a reviewer for a journal manuscript conducting panel analysis of IPEDS. Imagine that the author uses proportion of total FTE enrollments to allocate finance variables from the parent level to the child level, for systems where all finance variables are reported at the parent level. Or imagine that the author uses proportion of total current revenues to allocate parent-level asset variables to the child level, for systems where asset variables are reported at the parent level but other finance variables are reported at the child level. The choice of a particular allocation factor variable is inherently arbitrary, and the resulting allocation factors will not reflect the “true” allocation from parents to children. Readers—and reviewers in particular—are skeptical of arbitrary decisions. Therefore, in light of the considerable effort to implement allocation solutions and in light of the disappointing results of these efforts, we recommend that researchers avoid allocation solutions altogether.

Common Data Challenges

This section describes solutions to common data challenges inherent when creating a panel dataset using IPEDS data. Parent-child reporting plays an important role in most data challenges. We begin by discussing merging issues. In particular, we discuss merging data from different IPEDS survey components, merging IPEDS data to Office of Postsecondary Education and Office of Federal Student Aid data, and merging organization-level HEGIS/IPEDS data to NCES student-level data. Second, we discuss collapsing organization-level data into state-level data to conduct state-level analyses. Third, we discuss appending HEGIS and IPEDS data to create a HEGIS/IPEDS panel. Fourth, we discuss institutional closures and mergers. Fifth, we discuss key changes in IPEDS that affect many empirical applications of IPEDS, specifically change in accounting standards used in the Finance component, change in degree classification systems used by the Completions component, and change race/ethnicity categories that affect several IPEDS survey components.

Merging Data

Merging IPEDS Survey Components

Panel analyses of IPEDS data often require researchers to merge data from different IPEDS survey components, where the resulting analysis dataset has variables from multiple survey components and one observation per organization-year. “Merging” *joins* observations from the dataset currently in memory to observations from another dataset, so that the merged dataset contains variables from both datasets, as demonstrated in Table 11.16 which contains merged data from the Fall Enrollments and Finance survey components for the North Orange County Community College District. “Appending” *adds* observations from one dataset to observations from another dataset, so that the appended dataset contains observations from both input datasets. For example, the Fall Enrollment variables in Table 11.12 contain observations from Cypress College and Fullerton College for the years 2002–2007 because individual years of Fall Enrollment data—e.g., 2002, 2003, and 2004—were appended one to another.

When merging, observations from the dataset currently in memory are “matched” by one or more identification variables (e.g., institutional ID, student ID) to observations in the other dataset.²¹ In IPEDS, the UNITID variable matches observations

²¹ A “one-to-one” merge means that the “matching variable” uniquely identifies observations in both the current dataset and the using dataset. A “many-to-one” merge means that there may be duplicate observations of each matching variable in the current dataset but the matching variable uniquely identifies observations in the other dataset.

from one survey component to observations from another survey component (in HEGIS, the matching variable is called a FICE code). When merging a dataset containing a single year of IPEDS data from one survey component (e.g., Fall Enrollment data from 2002) to a dataset containing a single year of IPEDS data from another survey component (e.g., Finance data from 2002), researchers merge “by” UNITID. When merging a dataset containing multiple years of IPEDS data from one survey component (e.g., Fall Enrollment data from 2002 to 2007) to a dataset containing multiple years of IPEDS data from another survey component (e.g., Finance data from 2002 to 2007), researchers merge “by” UNITID and year.

Researchers can create a panel dataset from multiple IPEDS survey components in one of two ways. First, one could merge data separately for each year—e.g., merge finance variables from 2012 to Fall Enrollment variables to 2012 by UNITID and merge finance variables from 2013 to Fall Enrollment variables from 2013 by UNITID—and then append the merged datasets to one another. However, it is more efficient to follow a second approach of creating separate panel datasets for each survey component (e.g., append finance variables for 2012 and 2013 and append Fall Enrollment variables for 2012 and 2013), and then merging the two panel datasets by UNITID and year.

Time is a tricky element when merging different IPEDS survey components. Panel analyses of IPEDS usually measure time in academic years. For example, 2012 refers to the 2011–2012 academic year running from August 2011 through July 2012. When merging IPEDS survey components by UNITID and year, researchers must be careful that merged UNITID-year observations from each survey component represent the same academic year. For example, the IC and Completions survey components are collected each fall. IC data collected in the fall of 2012 measures Institutional Characteristics for the fall of 2012, which corresponds to the 2012–2013 academic year. However, Completions data collected in the fall of 2012 corresponds to the 2011–2012 academic year. Fall Enrollment and Finance survey components are collected each spring. Fall Enrollment data collected in the spring of 2013 corresponds to the 2012–2013 academic year, but Finance data collected in the spring of 2013 corresponds to the 2011–2012 academic year. Prior to merging data from different survey components, researchers should read about the data collection cycle on the “IPEDS Resource Center” website and should read the documentation associated with each data file to confirm what academic year the file refers to.

The primary complication in merging data from different survey components is the parent-child relationship. In the example of North Orange County Community College, Finance data were reported at the District Office (UNITID=120023) from 2002 to 2004 (Table 11.13). Fall Enrollment data were reported at the Fullerton College campus (UNITID=114859) and the Cypress College campus (UNITID=113236) from 2002 to 2004 (Table 11.9). Therefore, merging Finance data and Fall Enrollment data by UNITID and year would not work for 2002–2004. Researchers must collapse data from each survey component to the parent level prior to merging, as demonstrated in Tables 11.14 and 11.15. Once each survey component is collapsed to the parent level, merging is straightforward, as shown in

Table 11.16. However, even after collapsing all survey components to the parent level, merging data from two survey components will usually result in merges with missing data because some specific campuses may not fill out all survey components in each year.

It is inappropriate to collapse IC data to the parent level because IC variables (e.g., highest degree offered, cost of room and board) should not be summed across observations in a parent-child relationship. Therefore, researchers should generally drop “child” observations from IC data that do not merge with data from other survey components that have already been collapsed to the parent level.

Researchers may want to merge IPEDS data to the DCP Database to add variables that were not included in the latter. However, the parent-child relationship makes this difficult. Consider an observation from one year of data where the UNITID is a “parent” in the DCP Database. In the DCP Database that UNITID represents aggregated data from parent in child campuses, but in the raw IPEDS data, that UNITID only represents data for the parent institution. Therefore, it is often inappropriate to merge DCP data with raw IPEDS data. To correctly merge DCP data to raw IPEDS data, the researcher would have to collapse raw IPEDS data to the parent level using the DCP parent-child list. Unfortunately, this is not yet possible because DCP has not made their parent-child list publicly available.

Merging IPEDS to Office of Postsecondary Education/Office of Federal Student Aid Data

Some research questions may require researchers to create a panel dataset that incorporates variables from different data sources from the US Department of Education (e.g., IPEDS data and Office of Postsecondary Education/Office of Federal Student Aid data). For example, Deming et al. (2012) constructed a panel dataset from 1999–2000 to 2008–2009 using IPEDS data for enrollments (fall, 12-Month, and full-time equivalent), degrees and awards, tuition, revenues and expenditures, and other Institutional Characteristics. This data was matched to institution-level data of Pell Grants, student loan volumes, and cohort default rates from the Office of Federal Student Aid. In this article, the authors were trying to assess student outcomes for the for-profit institutions relative to other higher education institutions after adjusting for observable differences in students who have attended these different types of schools. Another example of a research question that might more easily enable merging different data sources is: “How many students receive grants and scholarships from colleges and how much do they receive?”

In this section, we first describe the general process for creating a panel data that combines different data sources from the US Department of Education. We then describe the major data-matching challenges that researchers may face when merging different data sources and how to overcome those challenges. The general process for creating a panel dataset that incorporates IPEDS data, Office of Postsecondary Education, and Office of Federal Student Aid data is, first, to create separate panel datasets from each data source (e.g., a panel dataset of IPEDS

Table 11.20 Pell grant volume recipient from OFSA, 2000–2011

Institution	Recipients	OPEID	Year
Pennsylvania State University (The)	15,230	00332900	2000
Pennsylvania State University (The)	15,108	00332900	2001
Pennsylvania State University (The)	15,611	00332900	2002
Pennsylvania State University (The)	16,042	00332900	2003
Pennsylvania State University (The)	15,947	00332900	2004
Pennsylvania State University (The)	15,884	00332900	2005
Pennsylvania State University (The)	15,068	00332900	2006
Pennsylvania State University (The)	15,692	00332900	2007
Pennsylvania State University (The)	16,633	00332900	2008
Pennsylvania State University (The)	20,467	00332900	2009
Pennsylvania State University (The)	23,574	00332900	2010
Pennsylvania State University (The)	23,934	00332900	2011

measures; a panel dataset of the Office of Postsecondary Education measures; and a panel dataset of the reports from the Office of Federal Student Aid measures). Next, merge each panel dataset by institutional identifier and year. The two main challenges that a researcher may face when matching IPEDS data with data from other data sources are (a) inconsistent school identifiers and (b) incomplete matches.

Inconsistent School Identifiers

The main problem that researchers will have to deal with when merging two different data sources such as IPEDS components and data from the Office of Federal Student Aid is that they use different identifier codes for institutions and campuses (Steele 2011). IPEDS uses a six-digit identifier called UNITID for each campus, whereas data from the Office of Federal Student Aid uses the eight-digit OPEID code that is a hierarchical eight-digit code discussed in the Program Participation Agreements section of this chapter.

To remedy this problem, researchers may use the IPEDS Institutional Characteristic (IC) files that include both the UNITID and the eight-digit OPEID. Researchers can first merge data from the Federal Student Aid Data Center with the Institutional Characteristic files from IPEDS using the eight-digit OPEID and then merge this file with any other IPEDS components using the UNITID. But this remedy is imperfect because almost all datasets from the Federal Student Aid Data Center have eight-digit OPEIDs where the two-digit branch code is always “00.” In other words, IPEDS data often includes observations for Title IV branch campuses (that have two-digit suffixes that do not equal “00”) but, from the Office of Federal Student Aid, do not have separate observations for Title IV main campus and the Title IV branch campuses.

As an example, Table 11.20 shows that Pennsylvania State University has only one observation per year in the Student Financial Aid files (OPEID=00332900). However, Table 11.5 shows that Pennsylvania State University has 24 observations

per year in IPEDS data, with each observation representing a different campus. The campus assigned with OPEID=00332900 in the IPEDS file is Pennsylvania State University at University Park. If the researcher merges the two datasets, he or she might end up using the total amount of financial aid data for Pennsylvania State University with enrollment or graduation data from only one campus. We call this mistake “incomplete matches.”

Incomplete Matches

To illustrate some of the challenges of and remedies for incomplete matches, we provide examples using the grant volume data from the Federal Student Aid Data Center and the 12-Month Enrollment component files available on the IPEDS Data Center.

As stated above, even if Student Financial Aid data has an eight-digit OPEID, a one-to-one merge to IPEDS data may not be correct. The main question that researchers need to ask is: How do I know if the merge is correct? There are several ways to ascertain the answer to this question, the first being when one obtains nonsensical results, for example, ending up with more Pell Grant recipients than there are enrollees as reported in the IPEDS 12-Month Enrollment file. To overcome such errors, first collapse IPEDS data at the OPEID eight-digit code level to the OPEID six-digit code level. For example, Table 11.21 displays several Pennsylvania State University campuses with different eight-digit OPEID codes but the same six-digit OPEID during the 2009–2011 period. Table 11.22 presents the same data after collapsing from the eight-digit OPEID level to the six-digit OPEID level. The last step is to merge the IPEDS data (Table 11.22) and the Federal Student Aid data (Table 11.20) by OPEID six-digit code and year as demonstrated in Table 11.23.

Merging HEGIS/IPEDS Data to NCES Student-Level Data

Researchers conducting student-level analyses using nationally representative NCES surveys may wish to merge organization-level variables from HEGIS or IPEDS to the student-level survey data (e.g., Chen 2012). NCES has created several nationally representative student-level surveys that track the progress of students from high school to postsecondary education and into the labor market. For example, the National Longitudinal Survey (NLS:72), the High School and Beyond Survey (HS&B:80), the National Education Longitudinal Survey (NELS:88), and the Education Longitudinal Survey (ELS:2002) track the high school senior classes of 1972, 1982, 1992, and 2004, respectively. NCES student-level datasets—especially for more recent surveys (e.g., ELS:2002)—include many postsecondary organization-level variables (e.g., average SAT score of the incoming class, tuition price). But if the desired organization-level variables do not exist in the student-level datasets, then the researcher may want to merge variables from HEGIS or IPEDS. NCES student-level surveys often contain transcript data, an institutional ID code, and the institutional name for each

Table 11.21 IPEDS data for Pennsylvania State University, selected campuses

UNITID	Institution	Undergrad full-time	Undergrad part-time	OPEID-8	OPEID-6	Year
214777	Pennsylvania State University-Main Campus	36,749	1,239	00332900	003329	2009
214777	Pennsylvania State University-Main Campus	37,485	1,145	00332900	003329	2010
214777	Pennsylvania State University-Main Campus	37,347	1,247	00332900	003329	2011
214616	Pennsylvania State University-College Of Medicine	0	0	00332901	003329	2009
214616	Pennsylvania State University-College Of Medicine	0	0	00332901	003329	2010
214616	Pennsylvania State University-College Of Medicine	0	0	00332901	003329	2011
214689	Pennsylvania State University-Penn State Altoona	3,778	235	00332903	003329	2009
214689	Pennsylvania State University-Penn State Altoona	3,924	223	00332903	003329	2010
214689	Pennsylvania State University-Penn State Altoona	3,888	240	00332903	003329	2011
214698	Pennsylvania State University-Penn State Beaver	696	140	00332904	003329	2009
214698	Pennsylvania State University-Penn State Beaver	658	193	00332904	003329	2010
214698	Pennsylvania State University-Penn State Beaver	690	216	00332904	003329	2011
214591	Pennsylvania State University-Penn State Erie-Behrend Coll	3,963	255	00332905	003329	2009
214591	Pennsylvania State University-Penn State Erie-Behrend Coll	3,988	302	00332905	003329	2010
214591	Pennsylvania State University-Penn State Erie-Behrend Coll	3,929	323	00332905	003329	2011
214810	Pennsylvania State University-Penn State Schuylkill	885	131	00332914	003329	2009
214810	Pennsylvania State University-Penn State Schuylkill	859	148	00332914	003329	2010
214810	Pennsylvania State University-Penn State Schuylkill	876	158	00332914	003329	2011
214652	Pennsylvania State University-Penn State Worthington Scr. an.	1,108	271	00332915	003329	2009
214652	Pennsylvania State University-Penn State Worthington Scr. an.	1,084	304	00332915	003329	2010
214652	Pennsylvania State University-Penn State Worthington Scr. an.	1,096	290	00332915	003329	2011
214713	Pennsylvania State University-Penn State Harrisburg	2,482	481	00332920	003329	2011
214731	Pennsylvania State University-Brandywine	1,420	195	00332921	003329	2009
214731	Pennsylvania State University-Brandywine	1,383	224	00332921	003329	2010
214731	Pennsylvania State University-Brandywine	1,376	237	00332921	003329	2011
212018	The Dickinson School Of Law Of The Pennsylvania State Uni.	0	0	00332922	003329	2009
212018	The Dickinson School Of Law Of The Pennsylvania State Uni.	0	0	00332922	003329	2010
212018	The Dickinson School Of Law Of The Pennsylvania State Uni.	0	0	00332922	003329	2011

Table 11.22 12-Month Enrollments for Pennsylvania State University, after collapse

UNITID	Year	Institution	Undergrad full-time	Undergrad part-time	OPEID-8	OPEID-6
214777	2000	Pennsylvania State University-Main Campus	1,397	57	00332900	003329
214777	2001	Pennsylvania State University-Main Campus	1,289	33	00332900	003329
214777	2002	Pennsylvania State University-Main Campus	1,399	39	00332900	003329
214777	2003	Pennsylvania State University-Main Campus	1,511	50	00332900	003329
214777	2004	Pennsylvania State University-Main Campus	1,472	58	00332900	003329
214777	2005	Pennsylvania State University-Main Campus	1,460	63	00332900	003329
214777	2006	Pennsylvania State University-Main Campus	1,652	62	00332900	003329
214777	2007	Pennsylvania State University-Penn State Main Campus	1,814	46	00332900	003329
214777	2008	Pennsylvania State University-Main Campus	1,863	41	00332900	003329
214777	2009	Pennsylvania State University-Main Campus	1,923	50	00332900	003329
214777	2010	Pennsylvania State University-Main Campus	2,001	34	00332900	003329
214777	2011	Pennsylvania State University-Main Campus	1,763	46	00332900	003329

Undergrad full-time: total of undergrad students full-time

Undergrad part-time: total of undergrad students part-time

Table 11.23 Merged IPEDS 12-Month Fall Enrollment and Pell Grant Recipients from Office of Financial Student Aid

UNITID	Year	Institution	Undergrad full-time	Undergrad part-time	OPEID-6	Recipients
214777	2000	Pennsylvania State University-Main Campus	1,397	57	003329	15,230
214777	2001	Pennsylvania State University-Main Campus	1,289	33	003329	15,108
214777	2002	Pennsylvania State University-Main Campus	1,399	39	003329	15,611
214777	2003	Pennsylvania State University-Main Campus	1,511	50	003329	16,042
214777	2004	Pennsylvania State University-Main Campus	1,472	58	003329	15,947
214777	2005	Pennsylvania State University-Main Campus	1,460	63	003329	15,884
214777	2006	Pennsylvania State University-Main Campus	1,652	62	003329	15,068
214777	2007	Pennsylvania State University-Penn State Main Campus	1,814	46	003329	15,692
214777	2008	Pennsylvania State University-Main Campus	1,863	41	003329	16,633
214777	2009	Pennsylvania State University-Main Campus	1,923	50	003329	20,467
214777	2010	Pennsylvania State University-Main Campus	2,001	34	003329	23,574
214777	2011	Pennsylvania State University-Main Campus	1,763	46	003329	23,934

postsecondary education institution the student attends. Student-level data from NLS:72 and HS&B:80 should be merged to the appropriate year of HEGIS data by FICE code using a many-to-one match (several students may attend the same institution). Student-level data from NELS:88 and ELS:2002 should be merged to the appropriate year of IPEDS data by UNITID also using a many-to-one match.

Unfortunately, earlier NCES surveys—especially NLS:72 and HS&B:80—often have the incorrect FICE codes for the postsecondary institution attended (e.g., sometimes providing the FICE code of a system office rather than the campus). Student-level observations with an incorrect FICE code may fail to match with the organization-level dataset or may match to the incorrect organization. Therefore, prior to merging student-level data to organization-level data, researchers must check that each FICE code in the student-level data is correct by comparing a list of institutional names and FICE codes from student-level data to a list of institutional names and FICE codes from organization-level data. Additionally, researchers must be careful that they record the FICE code for the appropriate campus of a multicampus organization/system since the Institutional Characteristics may differ greatly across them (e.g., University of Michigan–Ann Arbor is very different than the University of Michigan–Dearborn).

Collapsing Data to Create State- or Organization-Level Measures

Increasingly, policy researchers are conducting panel analyses where each observation represents a state-year (e.g., Arizona in 2008) rather than each observation representing an organization-year. For example, Toutkoushian and Hillman (2012) constructed a state-level panel dataset by merging state-level IPEDS measures by year and state to state-level measures from other data sources. They analyzed the effect of state-level expenditures on state appropriations (data from the “Grapevine” dataset), need-based grants (data from National Association of State Scholarship and Grant Aid Programs (NASSGAP)), and merit-based grants (data from NASSGAP); state-level higher education enrollments (from IPEDS); and state-level out-migration (also from IPEDS) (i.e., a state resident enrolling in an out-of-state postsecondary institution).

IPEDS data can also be used to create annual, state-level measures, such as total undergraduate enrollments per state and year, total number of bachelor’s degrees awarded per state and year, and total net tuition revenue per state and year, among others. The process for creating annual, state-level measures—e.g., total number of bachelor’s degrees awarded—follows these steps: first, sort the data by state and year; second, create a new variable that is the sum of bachelor’s degrees awarded at all institutions in each state and year; and third, keep the last observation for each state and year. The statistical programming for creating annual, state-level measures is almost identical to the statistical programming for collapsing child-level data to parent-level data.

When creating state-level measures from IPEDS data, it does not matter whether observations for public institutions represent one campus or all campuses in a state system. Therefore, the DCP Database can be used to create state-level IPEDS measures despite the fact that it collapses many state systems into a single observation. However, researchers creating state-level measures using raw IPEDS data should be careful about collapsing child-level data to the parent level for private nonprofit and private for-profit institutions because the parent may be located in a different state than the child. Therefore, state-level measures should generally be created prior to collapsing child-level data to the parent level. Finally, researchers must carefully consider the timing of data collection when merging state-level IPEDS measures to other state-level.

Appending HEGIS and IPEDS Data

Panel analyses that span HEGIS and IPEDS show that some variables we might consider to be “fixed”—e.g., highest degree awarded—may actually vary over time. For example, Jaquette (2013a) demonstrated that of the 678 organizations defined as “liberal arts colleges” by the 1973 Carnegie Classification, more half either no longer existed or “became a university” (defined by a change in organizational name) by 2010.

The HEGIS/IPEDS panel created to conduct this analysis was developed by appending HEGIS data to IPEDS data, separately for each survey component, and then merging panel data from different survey components by institutional ID code and year. As mentioned above, the FICE code is the ID variable for HEGIS data and UNITID is the ID variable for the IPEDS data. The greatest challenge in creating the HEGIS/IPEDS panel is defining an institutional ID variable that identifies each campus across all HEGIS and IPEDS years. Just as an IPEDS panel dataset has one observation per organization-year, with each distinct organization represented by a unique UNITID, a HEGIS/IPEDS panel dataset should have one observation per organization-year, with each distinct organization represented by a unique UNITID.

To facilitate the development of this panel, we created a FICE-UNITID crosswalk, which identified a unique UNITID for each FICE code. First, a preliminary FICE-UNITID crosswalk was created using IC IPEDS datasets from 1986 to 1999 because these datasets contain both UNITID and FICE code. The second step was to investigate why some FICE codes did not have a corresponding UNITID in the preliminary FICE-UNITID crosswalk. Investigations were conducted using targeted observation lists and web searches. We found two general reasons for the lack of correspondence: (1) the organization was no longer in existence (i.e., it closed or merged with another institution) prior to the inception of IPEDS; and (2) the FICE code represents a “child” campus which existed after the inception of IPEDS but never received a UNITID because it ceased reporting HEGIS data prior to the inception of IPEDS.

The third step in construction of this panel was to create a final FICE-UNITID crosswalk by assigning an ID code to FICE codes that did not have a UNITID.

Institutions that closed prior to IPEDS were assigned a UNITID equal to the FICE code, and branch campuses that ceased reporting HEGIS data prior to IPEDS were assigned a UNITID equal to the UNITID of the “parent” institution. To complete the crosswalk construction process, we merged the FICE-UNITID crosswalk to HEGIS data, separately for each survey component.²²

Once each observation of HEGIS data has a UNITID, it is fairly straightforward to create a HEGIS/IPEDS panel with data from multiple survey components. First, data from HEGIS years are appended to IPEDS years, separately for each survey component. Second, child-level observations are collapsed into parent-level observations, separately for each survey component. Finally, panel data from each survey component are merged by UNITID and year. However, the analysis period may often be shorter than the duration of a HEGIS/IPEDS panel because variables of interest may be unavailable in some years.

Institutional Closures and Mergers

As part of creating a HEGIS/IPEDS panel, we have investigated hundreds of closures and mergers, triangulating lists of HEGIS/IPEDS observations with published lists of closed/merged institutions, and web searches for each closed/merged institution.²³ When an organization closes, it ceases reporting IPEDS data and the UNITID of the closed organization is not used again. Our investigation of mergers

²² A complication to creating a FICE-UNITID crosswalk is that FICE codes were assigned to each campus with separate institution-level accreditation but were assigned to each campus that offered a complete degree program. Therefore, each FICE code may represent more than one UNITID. For example, the University of Nebraska–Lincoln (UNITID=X) and the University of Nebraska, Technical Agriculture (UNITID=X) have the same FICE code (FICE=X). If one FICE code is associated with more than one UNITID, then the FICE-UNITID crosswalk has duplicate observations for certain FICE codes. Merging HEGIS data (more than one observation per FICE code) by FICE code to a FICE-UNITID crosswalk with more than one observation per FICE code represents a “many-to-many” merge. This many-to-many merge would result in duplicate FICE-year observations for each FICE code associated with more than one UNITID. For example, for each year of HEGIS data, there would be two observations for the University of Nebraska–Lincoln. Therefore, the FICE-UNITID crosswalk should only contain one UNITID for each FICE code. For a FICE code associated with two UNITIDs, one UNITID will be connected to the FICE code, creating a panel that spans HEGIS and IPEDS years, and the other UNITID will not be connected to a FICE code, implying that the campus did not exist prior to IPEDS. Therefore, we recommend that researchers create a parent-child relationship when more than one UNITID refers to the same FICE code. The UNITID associated with the FICE code in the FICE-UNITID crosswalk should be defined as the parent, and the UNITIDs not associated with any FICE codes should be defined as children.

²³ All organizations eligible for Title IV eligibility must report IPEDS data. Therefore, IPEDS data can identify whether an organization loses Title IV eligibility, but IPEDS data alone cannot identify whether the organization ceased to exist. Furthermore, for multicampus organizations, IPEDS data alone cannot distinguish between a branch campus that closes versus a branch campus that ceases to report data at the child level.

suggests that one organization is typically the “acquiring” organization (e.g., DePaul University) and one organization is the “acquired” organization (e.g., Barat College). In all cases, the acquiring organization continues to report IPEDS data using its preexisting UNITID. In some cases, the acquired organization stops reporting IPEDS data immediately following the merger. More often, the acquired organization continues reporting IPEDS data under its preexisting UNITID for several years after the merger (e.g., Barat College).

How should researchers handle institutional mergers when creating IPEDS panels? “Acquired” organizations should be treated similarly to closed organization; once acquired, the organization—and its associated UNITID—should no longer exist. For the years prior to the merger, the UNITID of the acquiring organization should have no connection to the UNITID of the acquired organization. However, if the UNITID of the acquired organization continues to report data after the merger, then the researcher should replace the UNITID of the acquired organization with the UNITID of the acquiring organization for the post-merger years. This is equivalent to defining the acquired organization as a child of the acquiring organization in the post-merger years.

As described previously, the DCP Database handles many mergers inappropriately by defining the acquired organization as a child of the acquiring organization in both pre- and post-merger years. Therefore, the DCP Database collapses acquired organizations into acquiring organizations prior to the two organizations having any connection with one another.

IPEDS Changes: Accounting Standards, Degree Classification, and Race Categories

IPEDS has undergone many important changes since its inception. This section discusses three IPEDS changes that create difficulties for many empirical applications of IPEDS data. First, we discuss change over time in accounting standards used by the IPEDS Finance component, which makes it difficult to create variables that span the desired analysis period or that span different types of institutions (i.e., public, private nonprofit, and private for-profit). Second, we discuss change over time in degree classification systems, which affects research on degree production and the adoption of degree programs (e.g., Kraatz and Zajac 1996; Rojas 2006). Third, we discuss recent changes in race/ethnicity categories used by IPEDS survey components.

Accounting Standards

The IPEDS Finance component has undergone a series of revisions to its accounting standards, and these have important repercussions for researchers interested in creating panel data. Although a detailed discussion of change over time in variables is beyond the scope of this chapter, we briefly discuss these broad changes.

HEGIS Finance data were collected from 1968–1969 to 1985–1986, and IPEDS Finance data have been collected since 1986–1987. Recommendations from the American Institute of Certified Public Accountants (1974) led to the revised “Common Form” Finance component which was used by all institutions from 1974–1975 to 1995–1996 (U.S. Department of Education. NCES 2000). Beginning in 1996–1997, the Finance component was revised for private not-for-profit institutions due to changes in accounting standards released by the Financial Accounting Standards Board (FASB) (U.S. Department of Education. NCES 2000). Then, in 1999–2000, a shortened version of the FASB accounting form was created for for-profit institutions (Fuller 2011). Prior to 1999–2000, most for-profit institutions did not complete the Finance component; instead they completed the Consolidated Survey.

Beginning in 2001–2002, the Governmental Accounting Standards Board (GASB) Finance form was phased-in for public institutions in order to satisfy new accounting standards implemented by GASB (U.S. Department of Education. NCES 2009). Starting with the 2003–2004 Finance data, all public institutions were required to use the GASB Finance form. In 2008–2009, the GASB Finance form was modified to align information collected from public institutions using GASB standards with that from institutions using the FASB standards (U.S. Department of Education. NCES 2009).

There are two general differences between the newer Finance survey forms—FASB for private nonprofit, FASB for private for-profit, and GASB for public institutions—and the older Common Form (U.S. Department of Education. NCES 2009). First, the Common Form used accrual accounting and the newer Finance forms use cash accounting (U.S. Department of Education. NCES 2000). Accrual accounting differentiates “unrestricted” and “restricted” revenues. Unrestricted revenues (e.g., tuition revenue, state appropriations) can be used for any purpose and are recognized upon receipt. Restricted revenues (e.g., research grants, donations) can be used only for specific purposes and are recognized once the revenue has been expended (e.g., graduate student tuition paid by a research grant). Cash accounting does not recognize a difference between restricted and unrestricted revenues, so all revenues are recognized upon receipt. Second, the newer Finance survey forms require detailed reporting about assets, liabilities, and equity, but the Common Form did not (U.S. Department of Education. NCES 2012). For example, the GASB Finance form for public institutions includes these components: Statement of Net Assets (Part A), Summary of Changes in Net Assets (Part D), Debt and Assets (Part L), and Endowment Assets Data (Part H).

Researchers creating panel measures from Finance data often want to know whether a particular variable can span changes over time in accounting standards. In general, many of the revenue and expenditure variables on the newer Finance forms can also be created for data years using the Common Form because the transition from accrual to cash accounting does not affect expenditures and generally has only modest effects on revenues (U.S. Department of Education. NCES 2000). Researchers should note that endowment revenue and private grant revenue from the Common Form are not consistent with the newer Finance forms. However, many variables relating to assets, liabilities, and equity that can be created on the

newer Finance forms cannot be created using the Common Form simply because the Common Form did not request these survey items.

Several excellent resources exist for researchers interested in creating specific variables that span longitudinal and cross-sectional differences in accounting standards. For example, the US Department of Education's NCES (2000) provided a very helpful discussion of the underlying differences in accounting principles used by the Common Form and the FASB form for private nonprofit institutions. US Department of Education's NCES (2000) also contained a variable crosswalk between the Common Form and the FASB form and identified which variables were not comparable between the two survey forms. US Department of Education's NCES (2009) provided a broad overview between FASB and GASB accounting standards. Fuller (2011) identified change over time in specific survey items available in each Finance form. Finally, the Delta Cost Project (DCP) Database Mapping File and the Delta Data Dictionary are perhaps the best resource for constructing specific finance variables. Both spreadsheets can be downloaded from the DCP Database website.²⁴ The Database Mapping File identifies which specific variables—organized by the categories of revenues, expenditures, and balance sheet—can be created over time and across accounting standards and identifies the “input” variables used to create each “output” variable.

Classification of Instructional Programs (CIP)

Many empirical studies have used HEGIS and IPEDS data from the Completions component to analyze degree adoption (e.g., Jaquette 2013b; Kraatz and Zajac 1996; Rojas 2006) and degree production (e.g., Turner and Bowen 1990). The creation of panel measures of degree production or degree adoption requires an understanding of degree classification systems used by HEGIS and IPEDS.

Table 11.24 indicates change over time in the degree classification systems used by the Completions component. From 1965–1966 through 1970, the HEGIS Completions component used a system that is now commonly referred to as the “pre-HEGIS” system (Huff et al. 1970). The “HEGIS” degree classification system was used from 1970–1971 through 1981–1982. The 1980 Classification of Instructional Programs (CIP) system was applied to HEGIS completions data beginning in 1982–1983 (Malitz 1981). The CIP has been revised periodically to reflect the creation of new degree programs (e.g., new degree programs that emerged following the creation of the internet). The 1985 CIP revision was introduced in 1986–1987 (Malitz 1987). The 1990 CIP revision was introduced in 1991–1992 (Morgan et al. 1991). The 2000 CIP was introduced in 2002–2003 (U.S. Department of Education. NCES 2002). The 2010 CIP was introduced in 2010–2011 (U.S. Department of Education. NCES 2010). However, the underlying

²⁴ <http://nces.ed.gov/ipeds/deltacostproject/>

Table 11.24 Change over time in degree classification systems

System	Academic years
Pre-HEGIS System	1965–1966 to 1969–1970
HEGIS System	1970–1971 to 1981–1982
1980 Classification of Instructional Programs (CIP)	1982–1983 to 1985–1986
1985 Classification of Instructional Programs	1986–1987 to 1990–1991
1990 Classification of Instructional Programs	1991–1992 to 2001–2002
2000 Classification of Instructional Programs	2002–2003 to 2009–2010
2010 Classification of Instructional Programs	2010–2011 to present

logic of the CIP has remained constant since the inception of the CIP. The CIP User Site provides general and technical information about the CIP.²⁵

The Classification of Instructional Programs (CIP) was created to overcome three flaws in previous degree classification systems. First, the CIP is a hierarchical classification system. Each individual degree program has a six-digit CIP code (e.g., 13.0406 is higher education administration). These individual degree programs are categorized within four-digit series (e.g., 13.04 is educational administration) and then within two-digit series (e.g., 13 is education). Therefore, researchers can create measures of degree production or degree adoption at the degree level (i.e., six-digit CIP), at the broader four-digit CIP level, or at the broadest two-digit level. By contrast, previous degree classification systems were hierarchical only at a very broad level. For example, in the HEGIS degree classification system, all degree programs relating to education were numbered 800–899, but there was no hierarchical organization within education programs.

Second, the CIP covers all postsecondary education degree levels. Prior to the CIP, separate classification systems existed for “higher education” degrees (pre-HEGIS and HEGIS systems) and for sub-baccalaureate vocational degrees (called the “Handbook VI” classification system) (Malitz 1981). In theory, each six-digit CIP degree (e.g., higher education administration = 13.0406) could be awarded at any level (e.g., associate’s degree). Third, whereas previous degree classifications systems only contained degree names, the CIP has detailed descriptions for each degree code, leading to higher levels of accuracy when institutions classify degrees.

In practice, we have found that degrees at the six-digit CIP code level—and even the four-digit CIP code level—are often coded inconsistently across institutions and even over time within institutions. For example, many universities offer an MA program in business that would be described as a “general” MBA program by a higher education researchers and administrators. However, our investigations of IPEDS Completions data found that these general MBA were assigned several different six-digit CIP codes, for example, 52.0101 (business/commerce, general), 52.0201 (business administration and management, general), or 52.0299 (business administration, management and operations, other). Sometimes individual

²⁵ <http://nces.ed.gov/ipeds/cipcode/Default.aspx?y=55>

institutions changed the CIP code assigned to general MBA programs, even in years that did not span revisions to the CIP system. Therefore, researchers creating degree production measures at the six-digit and four-digit CIP code level should first investigate the variation in how specific degree programs have been coded by institutions.

Degree classification “crosswalks” link codes from older degree classification systems to codes from newer degree classification systems. Huff et al. (1970) provided a crosswalk between pre-HEGIS and HEGIS degree codes and the Resources page within the CIP website contains crosswalks between HEGIS and the 1985 CIP and also crosswalks between different iterations of the CIP.²⁶ Researchers must use degree classification crosswalks to create degree adoption and production measures that span changes in degree classification systems. In particular, research on degree adoption must differentiate whether the first observation of a particular CIP code at an institution represents the adoption of a new degree program by the institution or whether it represents a previously existing degree program that received a new CIP code due to a revision in the CIP.

Jaquette (2011) used the following rule to create measures of degree adoption: observing a new CIP code at an institution would not be counted as degree adoption in the first 2 years of a new iteration of the CIP (e.g., the 2010 CIP was introduced in 2010–2011) if that CIP code was not used in prior iterations of the CIP. For example, the four-digit CIP code 15.16 refers to “nanotechnology” and was first introduced in the 2010 CIP. Observing the CIP code 15.16 for the first time in 2010–2011 or 2011–2012 would not be counted as degree adoption, but observing this code for the first time in 2012–2013 would be counted as degree adoption.

Race/Ethnicity Categories

The US Department of Education’s NCES (2007) describes the adoption of new race categories for IPEDS, which has affected the Fall Enrollment, 12-Month Enrollment, Human Resources, and the Graduation Rates survey components. The old race categories were as follows: nonresident alien; race and ethnicity unknown; black, non-Hispanic; American Indian/Alaskan Native; Asian/Pacific Islander; Hispanic; and White, non-Hispanic. The new race categories are as follows: non-resident alien; race and ethnicity unknown; Hispanic of any race; American Indian or Alaska Native; Asian; Black or African American; Native Hawaiian or other Pacific Islander; White; and two or more races.

New race codes were optional during phase-in years and mandatory afterwards. During optional years some institutions used the old race/ethnicity categories and some institutions used the new categories. For the Fall Enrollments and Human Resources components, the new race codes were optional in the 2008–2009 and 2009–2010 academic years, and mandatory beginning in the 2010–2011 academic year. For the Completions, 12-Month Enrollment, and Graduation Rates

²⁶ <http://nces.ed.gov/ipeds/cipcode/resources.aspx?y=55>

components, the new race categories were optional in 2007–2008, 2008–2009, 2009–2010 and mandatory beginning in 2010–2011. Note that we refer to the academic year covered by the data collection, but the US Department of Education’s NCES (2007) referred to the year the data were collected.

During the optional years, IPEDS data from each survey component contains three kinds of race variables: old race categories variables, new race category variables, and derived variables. Variables for the old race categories have non-missing observations for institutions that used the old race categories. Variables for the new race categories have non-missing observations for institutions that used the new race categories. IPEDS also created a set of derived race variables—roughly equivalent to the old race categories—that are non-missing for all institutions.

Researchers may want to conduct analyses using an analysis period that spans the years prior to new race categories, years when new categories were optional, and years when new race categories were mandatory. Researchers cannot use data reporting the old race categories to create variables approximating the new race categories (e.g., “Native Hawaiian or other Pacific Islander”) because these categories were not collected prior to change in race/ethnicity categories.

Researchers can use data reporting the new race categories to create variables approximating the old race categories. For example, the old category of “Asian/Pacific Islander” can be approximated by adding the new categories of “Asian” and “Native Hawaiian or other Pacific Islander.” However, for variables that span old and new race categories, years that use the old categories will not be truly comparable to years that use the new race categories. This is primarily because the category “two or more races” exists only in the new race categories and students that would have identified themselves as belonging to one racial/ethnic group under the old race categories may categorize themselves as “two or more races” under the new race categories. This affects empirical research using an analysis period that spans the change in race categories. Consider research on the effect of nonresident enrollments on racial diversity. Declines in the number of Black and Hispanic students due to adopting the “two or more races” category may be falsely attributed to growth in nonresident enrollments unless the proper year indicator variables are included in the model.

Conclusion: What Research Questions Can Be Answered Using IPEDS and DCP Data?

IPEDS Data

Having identified core IPEDS concepts and having discussed solutions to common data challenges, we conclude the chapter by discussing the kinds of research questions that can be addressed using IPEDS data and data from the DCP Database. IPEDS is an organization-level survey that yields organization-level

measures. IPEDS is useful for research questions where the dependent variable(s) and independent variable(s) can be reasonably measured at the organization level. For example, several papers have analyzed the relationship between organization-level factors and the adoption of academic degrees (Kraatz and Zajac 1996; Rojas 2006). Other papers have analyzed the relationship between organization-level factors and change in organizational name (Jaquette 2013a; Morphey 2002). Ehrenberg and Zhang (2004) analyzed the effect of growth in the number of non-tenure-line faculty on organization-level graduation rates. Ehrenberg et al. (2006) analyzed the effect of growth in the number of merit scholarships on socioeconomic diversity. Curs and Jaquette (2013) analyzed the effect of nonresident enrollment growth on racial and socioeconomic diversity.

IPEDS data are also useful for research questions where the independent variable is a measure of the external environment (e.g., a policy change or an economic change) and the dependent variable can be reasonably measured at the organization level. For example, Hillman (2013) analyzed the effect of local unemployment rate on community college enrollments. Long (2004) analyzed the effect of state adoption of merit-aid programs on institutional tuition levels and room and board price. Zhang and Ness (2010) analyzed the effect of state merit-aid expenditure on institutional enrollments by type of institution.

Dependent variables in many IPEDS analyses are often student-level measures that are aggregated to the organization level. For example, Webber and Ehrenberg (2010) analyzed organization-level measures of graduation and persistence. It can be problematic, however, to use organization-level data to study student-level variables. For example, a researcher might want to know how much institutional aid is devoted to resident students versus nonresident students at public research universities. The IPEDS Finance survey identifies the total amount of expenditure on institutional financial aid, but does not break out institutional aid expenditure by different categories of students. Therefore, when the independent variable is an organization-level measure and the dependent variable is a student-level measure, researchers should consider conducting student-level analyses. Transforming organization-level variables to student-level data is appropriate and exemplified by Chen (2012) and Doyle (2010).

IPEDS data are sometime insufficient even when the measure of students is appropriately measured at the organization level. Consider the hypothesis, "growth in nonresident enrollments will negatively affect the number of students from underrepresented racial groups." This hypothesis implicitly assumes that students from underrepresented racial groups are unlikely to be nonresident students, but this assumption may be incorrect. To properly account for any such differences the researcher would want to be able to identify the race and state residency status of each student. Unfortunately, the IPEDS identifies the number of students by race and (every other year) IPEDS identifies the number of freshman by state of residence, but IPEDS does not identify the number of students by race/ethnicity and by state residency. In general, the student measures in IPEDS (e.g., enrollments, graduation) cannot categorize students beyond the dimensions of gender and race/ethnicity. These limitations would disappear if student measures in IPEDS were created

from a student-unit record database that tracked the progress of all students in postsecondary education.

Finally, IPEDS cannot be used to address research questions that require detailed information about organizational subunits. In particular, IPEDS is a poor data source for cost studies. Researchers often hope to use IPEDS Completions data and Finance data to calculate how much it costs to produce a degree. Instructional costs differ dramatically by field of study and level of study, but the Finance survey does not break out instructional costs by field of study or level of study. Therefore, the best-cost studies use suborganization-level data (e.g., Middaugh et al. 2003). A related research topic is the extent to which student subsidies differ across students or whether some students cross-subsidize other students. Winston (1999) used IPEDS Finance data to calculate institutional average subsidy per student. However, IPEDS cannot be used to examine variation in subsidy per student or cross-subsidization (e.g., the extent to which MA students cross-subsidize PhD students).

DCP Database

The DCP Database is based on IPEDS data and can be used to answer many of the same questions that could be answered by creating a panel dataset from raw IPEDS data. However, many research questions cannot be answered using the DCP Database because of the way child-level organizations were collapsed into parent-level organizations and because the DCP Database contains only a subset of all the IPEDS variables.

The DCP Database is appropriate for most organizational analyses of community colleges. Community college campuses that ever reported Finance data at the parent level (e.g., district, main campus) were collapsed to the parent level. However, this limitation should not undermine most organizational analyses of community colleges because only three state systems were collapsed to the parent level (Kentucky, Louisiana, and Indiana). Similarly, most analyses of private nonprofit institutions using the DCP Database will not be affected by the DCP decision to collapse many branch campuses into the main campus. Additionally, the DCP Database can also be used to analyze for-profit institutions because for-profit institutions were not collapsed to the parent level. Finally, the DCP Database can be used for state-level analyses because collapsing is not an issue when analyzing at the state level.

However, the DCP decision to collapse many state university systems (e.g., all Pennsylvania State University campuses) into a single observation prevents its use for many research questions. The DCP Database should not be used to answer research questions that involve comparisons of one kind of public university to another kind of public university (e.g., master's universities vs. research universities). It also should not be used to answer research questions that involve comparisons of public universities to private universities (e.g., public research universities vs. private research universities).

Additionally, the way the DCP deals with parent-child collapsing makes it impossible to analyze institutional closures or mergers using the DCP Database. Many institutions that closed or merged were collapsed into the institution that acquired the campus, even for years prior to the closure/merger, and this solution is not appropriate. Finally, a research question might require some variables that are included in the DCP Database and some variables that are available only in raw IPEDS data. But before raw IPEDS data can be merged to the DCP Database, it must be collapsed to the parent level, using the DCP parent-child list, but to date that list has not been released, making this task impossible.

Our goal in this chapter was both to introduce the reader to the complexities of constructing panel datasets from IPEDS data and to provide some guidance on how to set about this task. We discussed some of the main features and challenges presented by IPEDS data, such as institutional level data and documentation, the issue of change over time in the HEGIS and IPEDS sampling universes—as well as its impact on determining institutional parent-child relationships—and the complexity of merging data from different IPEDS survey components, among others. We also discuss solutions to common data challenges that researchers may face. Despite its challenges, the analysis of panel data can be a powerful tool in higher education research, with the potential to uncover historical trends that have heretofore eluded researchers. We hope that researchers willing to construct their own panel dataset with IPEDS data find this contribution useful. We also believe that interested researchers would benefit from reading this chapter along with reading Fuller (2011), since Fuller (2011) provided an excellent discussion on change over time in specific IPEDS survey items by survey component, and could be a good complement to start working with IPEDS panel datasets.

Acknowledgments We thank Stephen DesJardins and Stephen Porter for their encouragement and valuable comments. Nicholas Hillman and Bradley Curs also provided guidance about data sources and statistical modeling. We thank Karina Salazar, Amanda Parkman, and Andrew Blatter for helpful editing. We would also like to thank current and former IPEDS staff—in particular Tara Lawley, Jessica Shedd, Allison Bell, Samuel Barbett, Colleen Lenihan, and Craig Bowen—answering the dozens of emails and phone calls that helped us understand the principles and intricacies of the IPEDS data. Thanks also to the Association of Institutional Research for providing dissertation funding to create the original HEGIS/IPEDS panel dataset this chapter is based on and to the Spencer Foundation for providing funding to improve this panel dataset.

References

- AICPA. (1974). *Audits of colleges and universities*. New York: American Institute of Certified Public Accountants.
- Association for Institutional Research. (2003). AIR Alert # 16: Changes in IPEDS for institutions with multiple campuses *AIR Alert*. Tallahassee: Association for Institutional Research.
- Association for Institutional Research. (2013). *Report mapping (IC-Header, 2012–13)*. Tallahassee: Association for Institutional Research.
- Brint, S. G. (2002). Data on higher education in the United States: Are the existing resources adequate? *American Behavioral Scientist*, 45, 1493–1522.

- Brint, S., Mulligan, K., Rotondi, M. B., & Apkarian, J. (2013). *College and Universities 2000 Databases*. Available from University of California, Riverside. <http://www.higher-ed2000.ucr.edu/databases.html>
- Cellini, S. R., & Goldin, C. D. (2012). *Does federal student aid raise tuition? New evidence on for-profit colleges*. Cambridge, MA: National Bureau of Economic Research.
- Chen, R. (2012). Institutional characteristics and college student dropout risks: A multilevel event history analysis. *Research in Higher Education*, 53(5), 487–505. doi:10.1007/s11162-011-9241-4.
- Cheslock, J. J., & Gianneschi, M. (2008). Replacing state appropriations with alternative revenue sources: The case of voluntary support. *Journal of Higher Education*, 79(2), 208–229.
- Cheslock, J. J., & Rios-Aguilar, C. (2011). Multilevel analysis in higher education 1920 research: A multidisciplinary approach. *Higher Education: Handbook of Theory and Research*, 26, 85–123.
- Cohen, M. P. (1990). *Error analysis of estimates for less-than-two-year postsecondary education institutions*. Washington, DC: National Center for Education Statistics.
- Congressional Research Service. (2007). *Institutional eligibility for participation in Title IV student aid programs under the Higher Education Act: Background and reauthorization issues*. Washington, DC: Congressional Research Service.
- Council for Higher Education Accreditation. (2002). *The fundamentals of accreditation: What do you need to know?* Washington, DC: Council for Higher Education Accreditation.
- Curs, B. R., & Jaquette, O. (2013, April). *Does non-resident enrollment growth lead to declines in socioeconomic and racial diversity?* Paper presented at the 2013 American Educational Research Association, San Francisco, CA.
- Delta Cost Project. (2011). *Delta Cost Project documentation of IPEDS database and related products*. Washington, DC: Delta Cost Project.
- Delta Cost Project. (n.d.). *Listing of “grouped” institutions in the Delta Cost Project IPEDS database*. Washington, DC: Delta Cost Project.
- Deming, D. J., Goldin, C., & Katz, L. F. (2012). The for-profit postsecondary school sector: Nimble critters or agile predators? *Journal of Economic Perspectives*, 26(1), 139–164.
- Doyle, W. R. (2010). Changes in institutional aid, 1992–2003: The evolving role of merit aid. *Research in Higher Education*, 51(8), 789–810.
- Ehrenberg, R. G., & Zhang, L. (2004). *Do tenured and tenure-track faculty matter?* (NBER Working Paper 10695). Cambridge, MA: National Bureau of Economic Research.
- Ehrenberg, R. G., Zhang, L., & Levin, J. M. (2006). Crafting a class: The trade-off between merit scholarships and enrolling lower-income students. *Review of Higher Education*, 29(2), 195–211. doi:10.1353/rhe.2005.0082.
- Fuller, C. (2011). *The history and origins of survey items for the Integrated Postsecondary Education Data System*. U.S. Department of Education. Washington, DC: National Postsecondary Education Cooperative.
- Hillman, N. W. (2013). Community colleges and labor market conditions: How does enrollment demand change relative to local unemployment rates. *Research in Higher Education, Advance Online Publication*, 54(7), 765–780. doi:10.1007/s11162-013-9294-7.
- Huff, R. A., Chandler, M. O., & NCES. (1970). *A taxonomy of instructional programs in higher education*. Washington, DC: National Center for Educational Statistics.
- Hyatt, J. A., & Dickmeyer, N. (1980). *An analysis of the utility of HEGIS finance data in conducting institutional and higher education sector comparisons*. Washington, DC: National Association of College and University Business Officers.
- Jaquette, O. (2011). *In pursuit of revenue and prestige: The adoption and production of master’s degrees by U.S. colleges and universities*. Ph.D. dissertation, Center for the Study of Higher and Postsecondary Education, School of Education, University of Michigan, Ann Arbor.
- Jaquette, O. (2013a). Why do colleges become universities? Mission drift and the enrollment economy. *Research in Higher Education*, 54(5), 514–543.
- Jaquette, O. (2013b). In pursuit of revenue and prestige: the production of master’s degrees by U.S. universities.

- Jaquette, O., & Curs, B. R. (2013, March). *The effect of state appropriations on non-resident enrollments at public universities*. Paper presented at the 2013 Association for Education Finance and Policy, New Orleans, LA.
- Kraatz, M. S., & Zajac, E. J. (1996). Exploring the limits of the new institutionalism: The causes and consequences of illegitimate organizational change. *American Sociological Review*, 61(5), 812–836.
- Legal Information Institute. (2013). 34 CFR 668.14 – Program participation agreement. Retrieved June 19, 2013, from <http://www.law.cornell.edu/cfr/text/34/668.14>
- Long, B. T. (2004). How do financial aid policies affect colleges? The institutional impact of the Georgia HOPE scholarship. *Journal of Human Resources*, 39(4), 1045–1066.
- Malitz, G. S. (1981). *A classification of instructional programs (CIP)*. Washington, DC: Center for Education Statistics.
- Malitz, G. S. (1987). *A classification of instructional programs (CIP)*. Washington, DC: Center for Education Statistics.
- Middaugh, M. F., Graham, R., Shahid, A., Carroll, C. D., & NCES. (2003). *A study of higher education instructional expenditures the Delaware Study of Instructional Costs and Productivity*. Washington, DC: U.S. Dept. of Education, Institute of Education Sciences, National Center for Education Statistics.
- Morgan, R. L., Hunt, S. E., & Carpenter, J. M. (1991). *Classification of instructional programs* (1990th ed.). Washington, DC: National Center for Education Statistics.
- Morphew, C. C. (2002). “A rose by any other name”: Which colleges became universities. *Review 1984 of Higher Education*, 25(2), 207–223.
- National Science Foundation. (2013). *WebCASPAR integrated science and engineering resources data system*. Retrieved June, 2013. <https://webcaspar.nsf.gov>
- Pennsylvania State University. (2013). *Penn State campuses*. Retrieved June 22, 2013, from <http://www.psu.edu/academics/campuses>
- Rojas, F. (2006). Social movement tactics, organizational change and the spread of African-American studies. *Social Forces*, 84(4), 2147–2166.
- Slaughter, S., & Leslie, L. L. (1997). *Academic capitalism: Politics, policies, and the entrepreneurial university*. Baltimore: Johns Hopkins University Press.
- Steele, P. (2011). *Suggestions for improvements to the collection and dissemination of federal financial aid data*. U.S. Department of Education. Washington, DC: National Postsecondary Education Cooperative.
- The University of Texas System. (2013). *The University of Texas System*. Retrieved June 22, 2013, from <http://www.utsystem.edu/>
- Toutkoushian, R. K., & Hillman, N. W. (2012). The impact of state appropriations and grants on 1999 access to higher education and outmigration. *Review of Higher Education*, 36(1), 51–90. doi:10.1353/rhe.2012.0063.
- Turner, S. E., & Bowen, W. G. (1990). The flight from the arts and sciences: Trends in degrees conferred. *Science*, 250(4980), 517–521.
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (1999). *Integrated Postsecondary Education Data System (IPEDS): An improved system*. Washington, DC: NCES.
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2000). *IPEDS finance data comparisons under the 1997 financial accounting standards for private, not-for-profit institutes: A concept paper*. Washington, DC: NCES.
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2002). *Classification of instructional programs: 2000 edition*. Washington, DC: NCES.
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2005). *Integrated Postsecondary Education Data System data quality study*. Washington, DC: NCES.
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2007). *Changes to race/ethnicity reporting to ipeds*. Retrieved June 19, 2013, from http://nces.ed.gov/ipeds/news_room/ana_Changes_to_10_25_2007_169.asp

- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2009). *IPEDS Finance DATA FASB and GASB: What's the difference?* Retrieved February 27, 2009, from http://nces.ed.gov/ipeds/factsheets/fct_ipeds_finance_03072007_1.asp
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2010). *Introduction to the classification of instructional programs: 2010 edition (CIP-2010)*. Washington, DC: NCES.
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2012). *Updated finance reporting solutions for jointly audited institutions*. Retrieved June 21, 2013, from http://nces.ed.gov/ipeds/factsheets/fct_new_finance_03072007_2.asp
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2013a). *About IPEDS*. Retrieved June 13, 2013, from <http://nces.ed.gov/ipeds/about/>
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2013b). *IPEDS glossary*. Retrieved June 22, 2013, from <http://nces.ed.gov/ipeds/glossary/>
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2013c). *IPEDS survey components and data collection cycle*. Retrieved June 13, 2013, from http://nces.ed.gov/ipeds/resource/survey_components.asp
- U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. (2013d). *Report and recommendations from IPEDS Technical Review Panel #1: Data reporting structures and clustering*. Retrieved June 19, 2013, from http://nces.ed.gov/IPEDS/news_room/trp_Technical_Review_03132003_28.asp
- U.S. Dept. of Education, & NCES. (1976). *Higher education opening fall enrollment, 1976*. Washington, DC: U.S. Dept. of Education, NCES.
- UA South. (2013). *About UA South*. Retrieved June 22, 2013, from <http://uas.arizona.edu/about-ua-south>
- University of California. (2013). *University of California*. Retrieved June 22, 2013, from <http://www.universityofcalifornia.edu/>
- University of Wisconsin System. (2013). *Campuses*. Retrieved June 22, 2013, from <http://www.wisconsin.edu/campuses/>
- Webber, D. A., & Ehrenberg, R. G. (2010). Do expenditures other than instructional expenditures affect graduation and persistence rates in American higher education? *Economics of Education Review*, 29(6), 947–958. doi:10.1016/j.econedurev.2010.04.006.
- Winston, G. C. (1999). Subsidies, hierarchy and peers: The awkward economics of higher education. *Journal of Economic Perspectives*, 13(1), 13–36.
- Winston, G. C. (2003). Differentiation among US colleges and universities. *Review of Industrial Organization*, 24, 331–354.
- Zhang, L. (2010). The use of panel data models in higher education policy studies. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 25, pp. 307–349). New York: Springer.
- Zhang, L. A., & Ness, E. C. (2010). Does state merit-based aid stem brain drain? *Educational Evaluation and Policy Analysis*, 32(2), 143–165. doi:10.3102/0162373709359683.