

Chapter 8 Using National Datasets for Postsecondary Education Research

John H. Milam, Jr.

Introduction

With the ubiquitous World Wide Web, there is unprecedented, free access to a growing number of valuable datasets about postsecondary education. Almost every federal and state agency, association, and education-related organization has been forced to justify its existence and rethink its programs by making its data collections more readily accessible and meaningful to its constituents. This is a consequence, in part, of the fact that there have never been more data collected, whether mandated or voluntarily, in print and online. This process was once a very costly enterprise of survey design, sampling, mailing, scanning, data entry, analysis, adjudication, publication, and marketing. However, Web-enabled data access has transformed the collection and dissemination of all types of postsecondary information. Complex annual data collections, such as the Integrated Postsecondary Education Data Systems (IPEDS), that took several years to be released and published are now available on the Web just weeks after they close.

While learning to use these datasets, users should not be discouraged by the sheer complexity and number. The online tools for creating extracts and for analysis have never been easier to use and are undergoing dramatic improvements in meeting user needs. However, this chapter is simply a starting point in the process; and researchers should carefully explore the range of online resources that are documented.

Once obtained, the skills needed to explore any one dataset and find patterns in data are easily translated into working with the data dictionaries and the methodological concerns of others. Regardless of the dataset, users need to understand data elements and how they are defined. Users need to ask how the data can be meaningfully queried, grouped, sorted, aggregated, graphed, displayed, analyzed, and reported. And most importantly, users must know the data integrity of a dataset.

The first section of this chapter focuses on a review of the different ways to approach locating datasets. The discussion will be limited to large data collections which may be used for a variety of purposes and include a large sample size. The major data collections of the federal government will be reviewed in depth in section two, along with some widely recognized surveys by national postsecondary education associations and commercial vendors. After finding out about a dataset, a researcher must understand how to access it, which is addressed in section three. Once users begin to explore a data source in depth, there are other questions which must be addressed, which are documented in section four. A brief review of emerging trends in data collection and references follow.

This chapter closely parallels the work of the author in directing a project for the National Postsecondary Education Cooperative (NPEC) funded by the United States Department of Education's National Center for Education Statistics (NCES), called ANSWERS.

ANSWERS is a Website with a set of online tools for “Accessing National Surveys with Electronic Research Sources” (<http://nces.ed.gov/npec/answers>).

ANSWERS tools include a matrix of variables from national surveys, a question bank from the national sample surveys, a definition bank, references regarding survey development and using the national datasets, an inventory of national surveys, information about survey developers, a list of datasets with information about variables, and a list of subjects/topical areas by dataset. Special audience pages are designed to offer help in finding data to associations, federal agencies, state agencies, institutional researchers and planners, policy analysts, survey developers, and the media.

Figure 1: Screen capture of ANSWERS Website

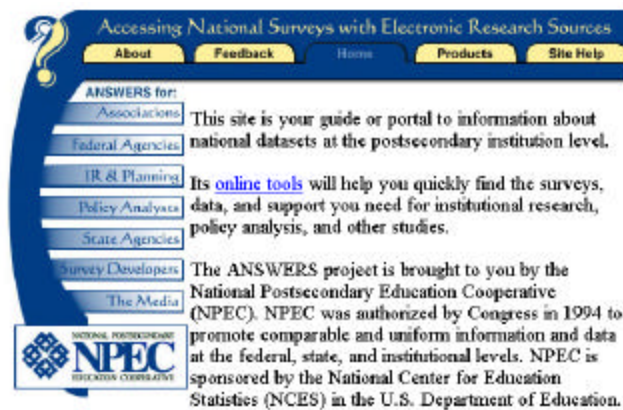


Table 1: Purposes of the NPEC ANSWERS Project

Purposes of the NPEC ANSWERS Project
<ul style="list-style-type: none"> • Describe and analyze information on frequently used surveys of postsecondary education institutions and national sample surveys • Promote better use of the national datasets • Provide a compendium of definitions and questions to promote best practices in survey development • Reduce data collection burden on institutions • Improve the comparability of data across surveys by promoting de facto standards

I. Using Different Lenses for Finding Data

In conducting institutional research, there are different approaches or lenses to use in finding needed data: topical, source, level of aggregation, collection method, time/date, and desired manipulation of the data.

The most obvious of these approaches is *topical*, in which the researcher looks for data by broad subject (*i.e.*, faculty, students, or financial aid) The search is then narrowed (*i.e.*, faculty rank or tenure status). The researcher must have knowledge of specific data-sets and/or access to tools that will assist in quickly finding the data needed by topic. One example of these tools is the matrix tool from ANSWERS which allows users to select a broad subject area, then to pick more narrow terms that are within that subject to find related data. A sample screenshot is shown below, illustrating the search for enrollment data that break out freshmen profile information in the IPEDS Fall Enrollment (EF) survey.

Figure 2: Search for enrollment data on freshmen profile

Matrix of Variables from National Surveys

Subjects:	Topics:	Datasets:	Variables:
Administration	(Enrollment)	(Freshmen Profile)	(IPEDS EF 01)
Admissions	Age	CDSE 00	1st-time, 1st-yr students graduated HS past 12 mo
Assessment	Certificate	College Board 00	1st-time first-year students (degree-seeking only)
Athletics	Classroom	IPEDS EF 00	State Residence of student when 1st admitted
Characteristics	Contact Hours	IPEDS EF 01	Status residence of first-time first-year students
Completions	Credit Hours	IPEDS EF 04	
Crime	Demographics	IPEDS EF 02	
Discipline	Extension	IPEDS EF 06	
Enrollment	First Professional	IPEDS EF 07	
Equipment	Foreign Countries	IPEDS EF 01	
Ethnicity	Freshmen Profile	IPEDS EF 04	
Facilities	FTE	IPEDS EF 04	
Faculty	Graduate	IPEDS EF 01	
Finance	Headcount	IPEDS EF 04	
Financial Aid	Level	IPEDS EF 01	
Gender	Minor	IPEDS EF 04	
Hospital	Postdocs		
Institution	Programs Offered		
Library	Remedial		
Military	Residence		
Public Service	Status		
Research	Summer		
Respondents	Transfer		
Salaries	Undergraduate		
Staff			
Student Changes			
Student Services			
Survey			

Another special tool for locating data by topic is the SHEEO Online Access to Resources (SOAR) Website, developed by the State Higher Education Executive Officers association. SOAR is “designed to support postsecondary education research and policy analysis” (<http://www.sheeo.org/soar/>).

SOAR allows users to search through two broad compendiums, “National Data Sources on Higher Education” and “Resources on Teacher Mobility.” In addition to locating sources of data by seven broad topics and over fifty keywords, SOAR describes each collection and provides links to key and related publications. The National Data Sources compendium is in its fifth edition and may be downloaded in PDF format. Its sections are

broken into general references, students and learning, faculty and staff, finance and facilities, K-12, adult/workforce, and other resources. Additional search options in SOAR allow users to specify an organization/agency, state, or level (institutional, international, national, regional, or state).¹

A second approach to locating data is by *source*. For example, a researcher may choose to investigate whether one of the federal agencies that collect postsecondary education-related data provides a dataset of interest. Most often, researchers use data and tools developed by NCES or the National Science Foundation (NSF). Many land grant institutions, and those with agricultural programs, also use data from the United States Department of Agriculture (USDA). There are many different types of online tools available from these agencies to facilitate the accessing and analyzing of their data.

The *level of aggregation* is a third lens used by researchers when locating data. At what level of aggregation does the researcher want to explore the data? Examples of levels of aggregation include state, region, Carnegie classification, individual institutions, and discipline. A sample survey about financial aid may not allow for analysis by state. Data from an institutional survey may be aggregated across schools to calculate a statewide figure, if the population of institutions is collected. Discipline-specific data may not always be available except in special studies or association surveys.

It is important for the researcher to separate the level of aggregation of the collection from the level of aggregation that the researcher is interested in for analysis. For example, in order to view tuition data by state, it is necessary to look at institutional data, since there is no collection of data from state offices.

Collection methods are a fourth lens to use in locating data. Are sample data adequate for the issue to be studied? Or are data needed for all individual institutions or departments or from a certain population? If sample data are to be used, is there sufficient stratification to make generalizations about the level of aggregation in which the user is interested? Even though a variable such as discipline is included in a survey, that does not necessarily mean that there is adequate weighting to perform every type of desired analysis. There are many other collection issues that must be addressed for a researcher to feel comfortable with the data. For example, how are missing data imputed and what is the response rate? These types of questions are further documented in section four.

Time, or date, is a fifth lens for locating data. Does the researcher want current data, three-year trends, historical data, or another view of the data with slices of time? It is important to document this need before looking for datasets. There are a number of very interesting longitudinal datasets available for looking at student cohorts, as well as science and engineering personnel, over time.

¹ The reader is referred to the author's AIR Professional File article entitled "Using the National Datasets for Faculty Studies" (Milam, 1999), for a detailed treatment of how data may be used for this topic. Other examples of references using the datasets are included in ANSWERS, in SOAR, and on the NCES and NSF Websites.

The sixth and final approach is determining *how the user wants to manipulate the data*. Sometimes, the user simply wants to download an extract for analysis locally using software tools such as SAS, SPSS, Access, or Excel. At other times, the data element dictionary and weighting scheme may be very complex. The user may then wish to use a software tool specific to the data. These more specific tools assist with merging multiple datasets over time, calculating new variables, developing appropriate weighting schemes for sample data, viewing frequencies by value label, and creating cross-tabs with descriptive statistics. While these can all be done with standard statistical software, some users prefer to use unique tools that will complete the entire task for them. Examples of unique tools include the Data Analysis Systems (DAS) from NCES and the WebCASPAR and SESTAT tools from NSF. These tools help users quickly understand a variety of collection issues.

In summary, when searching for data, it is important to review these six basic questions:

Table 2: Questions to ask in finding data

Questions to ask in finding data
<ul style="list-style-type: none"> • What are the topics of interest? • Is a certain data source preferred? • What level of aggregation will be used for the results? • Are sample or population data needed? • Which data are available based upon the date/time period of interest? • Is the dataset all that is necessary, or are special tools for analysis needed?

II. Understanding Major Data Collections

Due to their importance in the determination of national trends and policy analysis, the major data collections which are conducted by the primary federal agencies concerned with postsecondary education must be understood. Agencies conducting major data collections include the United States Department of Education, the National Science Foundation, and the United States Department of Agriculture. Other federal and state agencies, national associations, organizations concerned with postsecondary education, institutions, and commercial vendors also collect data. This section is not exhaustive. Rather, the highlights of collections are presented to provide a broad overview of what is available. The focus is on datasets that are available to users for free, as well as in electronic format suitable for further analyses.

U.S. Department of Education

NCES conducts the majority of institutional surveys and sample surveys for the U.S. Department of Education. Other offices are involved in efforts such as the Campus Crime/Security Survey, Equity in Athletics, and financial aid administration. The range of data collections may be grouped by population surveys of institutions, sample surveys of institutions and departments, and sample surveys of individuals (*i.e.*, faculty and students).

The Integrated Postsecondary Education Data System (IPEDS) surveys are at the heart of many NCES collection efforts. They are based on mandated reporting of over 9,000 institutions and include hundreds of data elements. IPEDS replaced the Higher Education General Information Survey (HEGIS), which collected data from 1965 to 1986 on accredited institutions. HEGIS included components for Earned Degrees/Completions, Finance, Residence and Migration, Salaries, Fall Enrollment, and Institutional Characteristics.

Where IPEDS surveys were submitted via paper survey forms from 1986 through 1999, there are now three combined data collections that are administered online at different times during the year (Fall, Winter, and Spring). In each dataset, references are still made to the print equivalent surveys, many of which have existed since HEGIS.

Discipline-specific data are collected by Classification of Instructional Programs (CIP) code, a taxonomy unique to NCES. A CIP 2000 version of this taxonomy has replaced the 1990 document, after extensive field review. Use of the new CIP taxonomy will be mandatory by 2004. Additional information about CIP 2000 is available at: <http://nces.ed.gov/ipeds/web2000/cip2000.asp>.

A complete list of NCES descriptions of these individual IPEDS surveys is included in Table 1. General information about IPEDS is available at: <http://nces.ed.gov/ipeds/index.html>.

Table 1 also includes the many sample surveys that have been conducted by NCES over time. Most of these are completed by persons, although survey records are supplemented by other Department of Education records such as financial aid information. Postsecondary Education Quick Information System (PEQIS) studies have examined such topics as distance learning, remedial education, and disabled students.

Some of the sample surveys involve gathering longitudinal data on a cohort of students over time, while other surveys are administered one-time only. Some of these collections have been discontinued, though the datasets are still made available for research. General information about the NCES postsecondary sample surveys is available at: <http://nces.ed.gov/surveys/SurveyGroups.asp?group=2>. More information is available from the Digest of Education Statistics 2001 appendix entitled "Guide to Sources," available at http://nces.ed.gov/pubs2002/digest2001/sources_1.asp.

At least three other major data collections at the institutional level are undertaken by the U.S. Department of Education, but are not managed by NCES. These include the Campus Crime and Security at Postsecondary Education Institutions Survey; reporting for the Equity in Athletics Disclosure Act (EADA); and the Campus-Based Programs Fiscal Operations Report and Application to Participate (FISAP), which is concerned with financial aid administration. Other collections of information are conducted by the Department for purposes of monitoring and compliance. The focus of this chapter, though, is on datasets available to the public for analysis.²

Table 3: Survey Datasets from NCES

TYPE	NAME	DESCRIPTION OF DATASET
Population – completed by school	IPEDS Institutional Characteristics (IC)	Institution names; addresses; telephone numbers; tuition, room and board charges; control or affiliation; calendar system; levels of degrees and awards offered; types of programs; selected student services; admission requirements; and accreditation status. IC surveys prior to 2000 collected instructional activity and unduplicated headcount data, which are now collected on the Enrollment survey.
Population – completed by school	IPEDS Fall Enrollment (EF)	Full- and part-time enrollment by racial/ethnic category and gender for undergraduate, first professional, and graduate student levels. Age distributions by student level and gender in odd numbered years. First time, first-year degree seeking student enrollments by home state of residence in even numbered years. Also in even-numbered years, four-year institutions are required to complete enrollment data by level, race/ethnicity, and gender for 9 selected disciplines for the Office for Civil Rights. In addition, the Enrollment survey now collects the instructional activity and unduplicated headcount data. Starting in 2002, unduplicated headcount by student level, and by race/ethnicity and gender of student are also be requested, as will total number of students in the entering class.
Population – completed by school	IPEDS Completions (C)	Numbers of associate's, bachelor's, master's, doctor's, and first professional degrees, and other formal awards, by field of study, and race/ethnicity and gender of recipient. Starting in 2001, completers of double majors by degree level, by race/ethnicity and gender of recipient, and by 6-digit CIP code will also be requested.
Population – completed by school	IPEDS Faculty Salaries (SA)	Full-time instructional faculty by rank, gender, tenure status, and length of contract; salaries and fringe benefits of fulltime instructional faculty. Data are collected annually, except for 2000.
Population – completed by school	IPEDS Fall Staff (S)	Number of institutional staff by occupational activity, full and part-time status, gender, and race/ethnicity. Data are collected in odd numbered years. Beginning with 1993, this survey replaced the EEO6 survey conducted by the Equal Employment Opportunity Commission.
Population – completed by school	IPEDS Employees by Assigned Position (EAP)	Headcount of full-time and part-time employees by faculty status and primary function/occupational activity. New survey, optional for 2001-02.
Population – completed by school	IPEDS Finance (F)	Current fund revenues by source (e.g., tuition and fees, government, gifts); current fund expenditures by function (e.g., instruction, research); assets and indebtedness; and scholarship and fellowship awards. Since 1997, Finance data have collected in different formats based on the institution's accounting standards (FASB or GASB).
Population – completed by school	IPEDS Student Financial Aid (SFA)	Started as the Institutional Price and Student Financial Aid study in 1999. Includes the number and percent of full-time, first-time, degree/certificate-seeking undergraduate students receiving student financial aid, by type of aid.
Population – completed by school	Academic Libraries (ALS)	Total library operating expenditures, full-time-equivalent library staff, service outlets, total volumes held at the end of the academic year, circulation, interlibrary loans, public service hours, patron count, reference transactions per typical week, and online services. Beginning in 1996, libraries were asked whether they offered the following electronic services: an electronic catalog that includes the library's holdings; electronic full-text periodicals; Internet access; library reference services by e-mail; and electronic document delivery to patron's account-address.
Population – completed by school	IPEDS Graduation Rates (GRS)	Number of students entering the institution as full-time, first-time, degree or certificate-seeking in a particular year (cohort), by race/ethnicity and gender; number completing within 150% of normal time to program completion; number transferred to other institutions; number of students receiving athletically-related student aid in the cohort and number completing within 150% of normal time.
Population – completed by schools	Equity in Athletics Disclosure Act (EADA)	EADA requires the Secretary of Education to collect information and provide to Congress a report on financial and statistical information on men's and women's collegiate sports, including athletic participation, staffing, coaching salaries, and revenues and expenses. Compara-

² As of this writing FISAP data are not yet available to the public, but this is the intention.

TYPE	NAME	DESCRIPTION OF DATASET
Population – completed by schools	Campus Crime and Security at Postsecondary Education Institutions Survey	ble to the NCAA report. Annual campus security report by institution with data on alleged criminal offenses reported to campus security authorities or local police agencies. The data collected do not necessarily reflect prosecutions or convictions for crime. Broken out by type of crime and location.
Population – completed by schools	Fiscal Operations Report and Application to Participate (FISAP)	Data on campus-based financial aid programs, by institution.
Sample – completed by school	Postsecondary Education Quick Information System (PEQIS)	The PEQIS was established in 1991 to collect issue-oriented data quickly and with minimum response burden. PEQIS was designed to meet the data needs of Department of Education analysts, planners and decision makers when information could not be collected quickly through traditional NCES surveys. The data collected through PEQIS are representative at the national level. The PEQIS employs a standing sample of 1,500 postsecondary institutions. Part of the National Household Education Surveys Program (NHES) data collection system. Household membership and individual characteristics; participation in adult education; type of program (vocational, occupational, basic skills, etc.); reasons for taking courses and barriers to participation in adult education. AES collected in 91, 95, 99, and 01. The Adult Education and Lifelong Learning survey (AELL-NHES:2001) also asked about less formal learning at work.
Sample – completed by person	Adult Education Survey (AES)	Part of the National Household Education Surveys Program (NHES) data collection system. Household membership and individual characteristics; participation in adult education; type of program (vocational, occupational, basic skills, etc.); reasons for taking courses and barriers to participation in adult education. AES collected in 91, 95, 99, and 01. The Adult Education and Lifelong Learning survey (AELL-NHES:2001) also asked about less formal learning at work.
Sample – completed by person	Recent College Graduates (RCG)	Date of graduation; field of study; graduates newly qualified to teach; further enrollment; financial aid; employment status (especially teacher employment characteristics); job characteristics and earnings; age; marital status; sex; and race/ethnicity. The survey was conducted in 1976, 1978, 1981, 1985, 1987 and 1991. More recent data collections of a longitudinal nature are undertaken by B&B.
Sample – completed by students, parents, & institutions	National Postsecondary Student Aid Study (NPSAS)	Includes data from three sources. From institutional student records: year in school; major field of study; type and control of institution; attendance status; tuition and fees; admission test scores; financial aid awards; cost of attendance; student budget information and expected family contribution for aided students; grade point average; age; date first enrolled. From student interviews, level; major field of study; financial aid at other schools attended during year; other sources of financial support; monthly expenses; reasons for selecting the school they are attending; current marital status; age; race/ethnicity; sex; highest degree expected; employment and income; community service. From parent survey: parents' marital status; age; highest level of education achieved; income; amount of financial support provided to children; types of financing used to pay child's educational expenses; occupation and industry.
Sample – see NPSAS	Baccalaureate and Beyond (B&B)	Follows an over-sample of graduating seniors from the National Postsecondary Student Aid Study. Includes information on education, employment, and other experiences. Collected current Federal aid and loan status information from ED records. B&B:93/94/97 included the collection of postsecondary transcripts. A new B&B cohort was created using NPSAS:2000 and this group was surveyed again in 2001 for the last time, similar to the RCG. B&B:93 will be studied again in 2003 with a focus on completion of graduate and professional school and educational indebtedness.
Sample – completed by students	Beginning Postsecondary Student Longitudinal Study (BPS)	BPS is designed specifically to collect data related to persistence in and completion of postsecondary education programs; relationships between work and education efforts; and the effect of postsecondary education on the lives of individuals. BPS has followed two cohorts. First time beginning students in a cohort from the NPSAS:90 were studied with follow-ups in 1992 and 1994. Beginners taken from NPSAS:96 were followed in 1998 and 2001. In addition to base-year NPSAS and all interview data, BPS: 96/98/2001 contains postsecondary entry test scores as well as financial aid records for the entire undergraduate period.
Sample – completed by faculty, chairs, & institutions	National Study of Postsecondary Faculty (NSOPF)	Includes data from three sources. From institutional survey: counts of faculty by rank; faculty hires and departures; tenure of faculty; tenure policies; retirement and other benefits for faculty. From department chair survey: faculty composition in department; tenure of faculty in department; tenure policies; rank; gender, and minority/non-minority status of faculty in department; faculty hires and departures in department; hiring practices; activities to assess faculty performance; professional and developmental activities. From faculty survey: socio-demographic characteristics; academic and professional background; field of instruction; employment history; current employment status including rank and tenure; outside employment; workload; courses taught; job satisfaction and attitudes; career and retirement plans; benefits and compensation.
Sample – completed by students	National Longitudinal Study of the H.S. Class of 1972 (NLS-72)	NLS-72 describes the transition of young adults from high school through postsecondary education and the workplace. This is considered the “grandmother” of the longitudinal studies conducted by NCES. Participants in the study were selected when they were seniors in high school in the spring of 1972, and in a supplementary sample drawn in 1973. The records include the "Base Year" survey; follow-up surveys in 1973, 1974, 1976, 1979, and 1986; high school records; and postsecondary transcripts (collected in 1984).
Sample – completed by students	National Education Longitudinal Study of 1988 (NELS:88)	Beginning with an 8th grade cohort in 1988, NELS:88 provides trend data about critical transitions experienced by young people as they develop, attend school, and embark on their careers. Data were collected from students and their parents, teachers, and high school prin-

TYPE	NAME	DESCRIPTION OF DATASET
Sample – completed by students	High School and Beyond (HS&B)	cipals and from existing school records such as high school transcripts. Cognitive tests (math, science, reading, and history) were administered during the base year (1988), first follow-up (1990), and second follow-up (1992). A third follow-up was conducted in 1994. All dropouts were retained in the study. The final follow-up of NELS:88 was conducted in 2000 and also includes the collection of postsecondary transcripts. This allows examination of the educational and labor market outcomes of the initial eighth-grade cohort of 1988 in the year 2000, when the majority of the cohort was 26 years old. HS&B describes the activities of seniors and sophomores as they progressed through high school, postsecondary education, and into the workplace. The data cover the period 1980 through 1992 and include parent, teacher, and high school transcript data, student financial aid records, and college transcripts in addition to student questionnaires. The HS&B survey included two cohorts: the 1980 senior class, and the 1980 sophomore class. Both cohorts were surveyed every two years through 1986, and the 1980 sophomore class was also surveyed again in 1992.

National Science Foundation

NSF supports the collection of a number of postsecondary-related datasets, outlined in Table 4. Data are collected from surveys of institutions, departments and programs, federal agencies, and persons. The primary focus is on information about science and engineering, though some surveys such as the biennial Survey of Doctorate Recipients (SDR) have been expanded at times to include humanities disciplines.³

The SDR, National Survey of Recent College Graduates (NSRCG) and the National Survey of College Graduates (NSCG) make up a larger, integrated, biennial collection called the Scientist and Engineer Statistics (SESTAT) data system. Another sample survey, the Survey of Public Attitudes Toward and Understanding of Science and Technology, looks at the public's knowledge and perception of science and engineering including educational data. General information about NSF data collections is available at: <http://www.nsf.gov/sbe/srs/survey.htm>.

Sample surveys are designed to provide national estimates for certain groups or activities. There are a number of publications driven by their production. NSF surveys have a unique taxonomy of disciplines, often in a finer level of detail than CIP codes for science and engineering fields. These come from the Survey of Earned Doctorates (SED) data on field of degree. The field or discipline of work documented in the SESTAT surveys are at a much less fine level of detail and match the occupation codes use by the U.S. Census Bureau. When working with disciplinary data, users will have to understand the cross-walk between the dataset they are using and their own departments or programs, which may be unique combinations of CIP or NSF taxonomies.

NSF's datasets place heavy emphasis on understanding the characteristics of doctoral scientists and engineers in the United States. These survey samples are taken from the Doctorate Records File (DRF), a database currently maintained by the National Opinion Research Center of the University of Chicago under contract to NSF. The DRF dates back to 1920, with the collection of graduation announcements and lists. The SED was begun in 1958 to continue this data collection. The SED is administered through graduate school deans and completed by all graduates with Ph.D., Sc.D., Ed.D., Doctor of Arts,

³ Note that NSF's science and engineering disciplines includes social sciences.

and other types of doctoral degrees. Professional school doctorates such as medicine, dentistry, and veterinary medicine are not included. There are approximately 1.4 million records in the 1920-2000 DRF.

Data on science and engineering programs by discipline at individual institutions are collected in the Survey of Graduate Students and Postdoctorates in Science and Engineering (GSS), the Survey of Research and Development Expenditures at Universities and Colleges (Academic R&D), and the Survey of Scientific and Engineering Research Facilities (R&D Facilities). Federal agencies submit data regarding specific postsecondary education institutions to NSF through two collections: the Survey of Federal Science and Engineering Support to Universities, Colleges, & Nonprofit Institutions (Federal Support); and the Survey of Federal Funds for Research and Development (Federal Funds).

Table 4: Survey Datasets from NSF

TYPE	NAME	DESCRIPTION OF DATASET
Population – completed by schools	Graduate Students and Postdoctorates in Science and Engineering (GSS)	Counts of graduate students by academic institution, department or program, geographic location, highest degree granted by institution (doctorate/master's), institutional control (public versus private), enrollment status, level of study (first year, beyond first year), sex, race/ethnicity, citizenship, primary source of financial support (e.g., NSF, NIH, etc.) of full-time students, primary mechanism of financial support (e.g., fellowship, research assistantship, etc.) of full-time students. Counts of postdoctoral fellows and doctoral non-faculty research staff by academic institution, geographic location, highest degree granted by institution, institutional control, sex, and for postdoctorates, source of support and citizenship.
Population – completed by grads	Survey of Earned Doctorates Awarded in the United States (SED)	Sex; age; race/ethnicity; marital status; citizenship; disabilities; dependents; specialty field of doctorate; all institutions attended from high school to completion of doctorate; time spent in completion of doctorate; source of financial support for graduate study; education debt incurred; postdoctoral plans; educational attainment of parents.
Population – completed by federal agencies	Survey of Federal Funds for Research and Development (FFR&D)	Character of work (basic research, applied research, and development), Federal agency, Federal funds for research and development, Federal obligations (defined by initiating agency), Federal outlays for research and development, Federally-funded research and development centers (FFRDC), Field of science and engineering, Geographic location (within U.S.), Performer (type of organization doing work, i.e., intramural, extramural), R&D plant.
Population – completed by federal agencies	Survey of Federal Science and Engineering Support to Universities, Colleges, & Nonprofit Institutions	Academic institution, Federal agency, Geographic location (within U.S.), Highest degree granted, Historically Black Colleges and Universities, Obligations (defined by obligating agency), R&D plant, Type of academic institution (Historically Black Institutions/others), Type of activity (e.g., research and development; S&E instructional facilities), Type of institutional control
Population – completed by schools	Survey of Research and Development Expenditures at Universities and Colleges (Academic R&D)	Academic institution, Character of work (basic research, applied research, and development), Equipment Expenditures, Expenditures for S&E R&D, FFRDCs, Field of science and engineering (for Total and Federal sources only), Geographic location (within U.S.), Highest degree granted, Passed through funds (through the institution to subrecipients and also institution received as a subrecipient), Source of funds (Federal, State and local, industry, institutional, and other), Type of academic institution (doctorate-granting versus, non-doctorate-granting), Type of academic institution (Historically Black Institutions versus others), Type of institutional control
Sample – completed by persons w/ S&E doctorates	Survey of Doctorate Recipients (SDR)	Citizenship, Country of birth, Country of citizenship, Year of birth, Disability status, Educational history, Employed status (part-time, full-time), Employer size, Faculty rank, Field of degree(s), Geographic place of employment, Labor force status (employed, unemployed, not in labor force), Level of degree(s), Marital status, Number and age of children, Occupation, Second job, Postdoc indicator, Primary work activity (e.g., teaching, basic research, etc.), Race/ethnicity, Salary, Hours/weeks worked, Previous year earned income, Professional association membership, School enrollment status, Sector of employment (academia, industry, government), Sex, Tenure status.
Sample – completed by persons with S&E bachelor & master's	National Survey of Recent College Graduates (NSRCG)	Citizenship, Country of birth, Country of citizenship, Year of birth, Disability status, Educational history, Employed status (part-time, full-time), Employer size, Field of degree(s), Geographic place of employment, Labor force status (employed, unemployed, not in labor force), Level of degree(s), Marital status, Number and age of children, Occupation, Second job, Primary work activity (e.g., teaching, basic research, etc.), Race/ethnicity, Salary, Hours/weeks worked, Previous year earned income, Professional association membership,

TYPE	NAME	DESCRIPTION OF DATASET
degrees		School enrollment status, Sector of employment (academia, industry, government), Sex, GPA, Undergraduate debt.
Sample – completed by persons w/ at least a bachelors degree	National Survey of College Graduates (NSCG)	Citizenship, Country of birth, Country of citizenship, Age, Disability status, Educational history, Employed status (part-time, full-time), Employer size, Field of degree(s), Geographic place of employment, Labor force status (employed, unemployed, not in labor force), Level of degree(s), Marital status, Number and age of children, Occupation, Second job, Primary work activity (e.g., teaching, basic research, etc.), Race/ethnicity, Salary, Hours/weeks worked, Previous year earned income, Professional association membership, School enrollment status, Sector of employment (academia, industry, government), Sex.
Sample – completed by schools	Survey of Science and Engineering Research Facilities	Amount of S&E research space; adequacy of the amount of S&E research space; condition of S&E research space; completion costs and NASF of repair/renovation and construction of S&E research space; source of funding for repair/renovation and construction of S&E research space; planned repair/renovation and construction of S&E research space; deferred repair/renovation and construction of S&E research space. Other non-survey variables: whether institution awards doctorate degrees in S&E, HBCU status, whether public or private institution.
Sample – completed by persons	Survey of Public Attitudes	Acceptance of science and technology, Admiration of science and technology, Age, Attitudes towards science and technology policy areas, Educational level, Geographic location (within U.S.), Interest in science and technology, Occupation, Perceived impacts of science and technology, Personal activities regarding science and technology, Public knowledge about science and technology, Race/ethnicity, Sex.

U.S. Department of Agriculture

The office of Higher Education Programs (HEP) is part of the Cooperative State Research, Education, and Extension Service of the United States Department of Agriculture. As one of its many activities, HEP maintains the Food and Agricultural Education Information System (FAEIS). FAEIS provides statistical information about postsecondary education related to the food and agricultural sciences. These disciplines include agriculture, forestry, renewable natural resources, family and consumer sciences, veterinary medicine, and closely allied fields.

Some of the data for FAEIS is secondary, compiled from the collections of professional associations, such as the American Association of Veterinary Medical Colleges. HEP also collects survey data of its own under two disciplinary clusters: (1) Agriculture, Renewable Natural Resources, and Forestry; and (2) Family and Consumer Sciences. Data are collected about enrollment, degrees awarded, faculty, placement, and institutional information. General information about these collections is available at: <http://faeis.tamu.edu>.

Table 5 documents these HEP datasets.

Table 5: Survey Datasets from USDA

TYPE	NAME	DESCRIPTION OF DATASET
Population – completed by schools	Agriculture, Renewable Natural Resources and Forestry - Enrollment	Available for Two-Year, Baccalaureate, Masters and Doctoral enrollment in Baccalaureate granting Colleges of Agriculture, Renewable Natural Resources and Forestry. Historic information is available from 1984. Data are summarized by the following categories: institution, academic area, gender, minority, land grant affiliation (1862, 1890, Non-Land Grant), professional, and NASULGC defined region.
Population – completed by schools	Agriculture, Renewable Natural Resources and Forestry - Degrees Awarded	Degrees conferred information is available for Two-Year, Baccalaureate, Masters and Doctoral enrollment in Baccalaureate granting Colleges of Agriculture, Renewable Natural Resources and Forestry. Historic information is available from 1984. Data are summarized by the following categories: institution, academic area, gender, minority, affiliation (1862, 1890, Non-Land Grant), professional, and NASULGC defined region.
Population – completed	Agriculture, Renewable Natural Re-	Information summarizes the placement of Baccalaureate, Masters and Doctoral graduates in the agricultural, renewable natural resource and forestry sciences. Information is provided

TYPE	NAME	DESCRIPTION OF DATASET
by schools	sources and Forestry - Placement	using academic area of degree granted and placement cluster. Placement clusters identified were based on Bureau of Labor Statistics area of work categories. Average starting salaries for graduates placed is also summarized by academic area of degree granted.
Population – completed by schools	Agriculture, Renewable Natural Resources and Forestry - Faculty	Provided for resident instruction faculty only. New faculty hired information is available annually. Annual salary information by broad academic area is available annually. Comprehensive information (collected every five years) summarizes a variety of demographic characteristics for resident instruction faculty in Colleges of Agriculture, Renewable Natural Resources and Forestry.
Population – completed by schools	Agriculture, Renewable Natural Resources and Forestry - Institutional Information	Attributes for all degree granting (Baccalaureate and higher) institutions in the agricultural, renewable natural resource and forestry sciences. Names, addresses, phone, fax, and email information are available for administrative contacts in Colleges of Agriculture, Renewable Natural Resources and Forestry (Dean, Associate Dean and Assistant Dean levels). Information for these files is assimilated from information reported to FAEIS on an annual basis through professional association directories.
Population – completed by schools	Family and Consumer Sciences - Enrollment	Enrollment information is available for Two-Year, Baccalaureate, Masters and Doctoral enrollment in Baccalaureate granting Colleges/Schools/Units of Family and Consumer Sciences. This survey is initiated biennially. Data are summarized by the following categories: academic area, gender, minority status and professional association.
Population – completed by schools	Family and Consumer Sciences – Degrees Awarded	Degrees conferred information is available for Two-Year, Baccalaureate, Masters and Doctoral enrollment in Baccalaureate granting Colleges/Schools/Units of Family and Consumer Sciences. This survey is initiated biennially. Data are summarized by the following categories: academic area, gender, minority and professional association.
Population – completed by schools	Family and Consumer Sciences - Faculty	Provided for resident instruction faculty only. New faculty hired information is available annually. Current faculty information (collected annually) summarizes a variety of demographic characteristics for resident instruction faculty in Colleges/Schools/Units of Family and Consumer Sciences.
Population – completed by schools	Family and Consumer Sciences - Institutional Information	Names, addresses, phone, fax, email information are available for contacts in Colleges/Schools/Units of Family and Consumer Sciences. Information for these files is assimilated from information reported to FAEIS on an annual basis through professional association directories.

Other Major Data Collections

There are many hundreds of postsecondary-related surveys and data collections. A much smaller number of these are accepted as de facto standards that may be relied upon for quality data. Some of these quality datasets originate from national associations, such as the American Association of University Professors (AAUP). They are targeted to specific audiences within the postsecondary education community or address specific policy issues, such as faculty salaries, and studies of tenure and rank.

The collection of postsecondary education data is a very viable commercial enterprise. A number of vendors, from John Minter Associates to *U.S. News and World Report*, collect, repackage, and sell data in various formats for different readers and data users. The quality and accuracy of these commercial collections has increased considerably as a result of the Common Data Set Initiative (CDS), a collaborative effort among data providers and publishers.

According to the CDS Website, “The CDS is a set of standards and definitions of data items rather than a survey instrument or set of data represented in a database. Each of the postsecondary education surveys conducted by the participating publishers incorporates items from the CDS as well as unique items proprietary to each publisher.”

The primary CDS vendors have included the College Board, Peterson's, *U.S. News & World Report*, and Wintergreen/Orchard House. Other CDS Advisory Board members include representatives from the following associations: the American Association of

Community Colleges (AACC); the Association of American Collegiate Registrars and Admission Officers (AACRAO); the Association for Institutional Research (AIR); the National Association for College Admission Counseling (NACAC); the National Association of College and University Business Officers (NACUBO); the National Association of Independent Colleges and Universities (NAICU); and the National Association of Student Financial Aid Administration (NASFAA). Wintergreen/Orchard House has since withdrawn from this effort. Additional information is available at: <http://www.commondataset.org/>.

In using a dataset, the reliability and integrity of the collection must be assessed. If the survey developer incorporates CDS standards and definitions, along with those maintained by NCES and NSF, there is an obvious degree of quality. Users should be cautious about using datasets which purport to collect new types of interesting information but fail to maintain effective practices of survey design, such as the use of standard census dates and definitions and building on the integrity of existing federal collections.

The following table lists a sampling of these reliable datasets. The listing is not exhaustive and is not a statement about the quality of other datasets that are not listed. Rather, it was prepared initially as part of the NPEC ANSWERS project. The datasets are available for free or for sale at low cost in electronic format and are recognized as adhering to de facto standards and definitions.

Table 6: A Sample of Other National Datasets

AGENCY	SURVEY	DESCRIPTION OF DATASET
American Association of University Professors (AAUP)	Annual Faculty Compensation Survey [AAUP-FCS]	Faculty salaries, benefits, and tenure status by institution, rank, and gender
Association of Post-secondary education Facilities Officers (APPA)	Comparative Costs and Staffing Report for Educational Facilities	Facilities operations, energy & utilities, contracting/outsourcing, maintenance staffing, custodial and grounds operations, etc.
Association of Research Libraries (ARL)	ARL Statistics Questionnaire	Collections, expenditures, personnel, and public services
Chronicle of Higher Education	Facts & Figures. Almanac.	Miscellaneous institutional data from various sources, some collected by the <i>Chronicle</i> . Data include, but are not limited to, information about campus crime, faculty salaries, pay and benefits of college presidents, endowments, fundraising, research library holdings, government grants, gender equity in athletics, graduation rates, stipends and benefits for graduate students, student enrollment and characteristics, and tuition and fees.
College and University Personnel Association (CUPA)	Administrative Compensation Survey	Salaries and compensation data for administrators.
College and University Personnel Association (CUPA)	Mid-Level Administrative/Professional Salary Survey	Salaries and compensation data for mid-level administrative and professional personnel
College and University Personnel Association (CUPA)	National Faculty Salary Survey By Discipline and Rank	Separate public and private institution versions of this survey of faculty salaries by discipline and rank. Information on new assistant professor salaries.
College Board, The (CB)	Annual Survey of Colleges	Admissions; enrollment; transfer; expenses; financial aid; curriculum; international; faculty; graduate programs.
Council for Aid to Education (CAE)	Voluntary Support of Education	Gift income broken down by source, purpose and outright vs. deferred, with additional detail on numbers of donors and forms of giving.
Council of Graduate Schools (CGS)	CGS/GRE Survey of Graduate Enrollment	Enrollment, applications, and degrees
National Research Council (NRC)	Research-Doctorate Programs in the United States	Information on 3,634 programs in 41 fields at 274 universities that participated in a comprehensive study of the research-doctorate enterprise. Behind

		these data is a much larger information set containing details on 88,000 faculty members; 1 million publication, citation, and research grant records; and 16,000 questionnaires for the reputational survey of faculty quality and program effectiveness.
Peterson's	Annual Survey of Graduate and Professional Institutions and Units	General institutional information, enrollment, faculty, research affiliations and projects, library and computer facilities, expenses, housing, financial aid/support, student services, degree programs, application and acceptance, entrance and degree requirements, degrees awarded, and contacts.
Peterson's	Annual Survey of Undergraduate Institutions	General institutional information, enrollment, persistence, freshman admissions, academics (e.g., faculty counts by status and gender; special programs; graduation requirements; etc.), facilities and services, expenses (e.g., full- and part-time tuition and fees and room and board; tuition payment plans and waivers), campus life, undergraduate majors; contacts, etc.
U.S. News	Rankings and Guides	Undergraduate, graduate, and professional school rankings and guides with extensive data collection.
University of Oklahoma	Consortium for Student Retention Data Exchange Survey	Enrollment and retention data.

III. Getting Access to Datasets

Once a dataset of interest has been located, there are several steps to consider in evaluating whether or not it will meet a user's needs. It is particularly useful to look at a copy of the actual survey instrument. The ANSWERS Website includes links to each survey instrument and dataset; this is a good place to start. NCES, NSF, and USDA all provide links to their surveys. When using a dataset for the first time, it is helpful to match the data to those submitted by a specific institution using the same instrument. While the data dictionary is usually very clear, with a detailed description of each field and its possible value labels, there is no substitute for the importance of recognizing data that are in the right place on a form and in a dataset.

General information about each dataset should be reviewed, such as the number and type of respondents, year of administration, and how the data are made available. If the user does not want to analyze data using SAS, SPSS, or another statistical analysis software package, or needs a specialized query tool that is tailored to the dataset, it is much better to determine this at the outset.

IPEDS datasets are available in fixed, ASCII text format, with instructions for importing the data into Microsoft Access or Excel and for importing the data into SAS or SPSS with a "read" program. Those with multiple records per institution require more manipulation. As part of the DAS software and separately, with its own installed software, Electronic Code Books (ECB) are available for each of the sample surveys, as well as for IPEDS. The most current IPEDS ECB is 2000, which is based on the most recent adjudicated, official, and final release IPEDS data. These ECBs let users browse, select, and view data elements, including the variable name, label, description, and value labels. Frequencies are also provided, along with minimum, maximum, mean, and standard deviation for continuous data elements. Whenever possible, users should obtain the codebook, or data element dictionary, for a dataset, including value labels and frequencies. Users may want to download ECBs from the NCES site to learn more about a dataset. However, if the DAS is being used for sample surveys, this is a duplication of the functionality, since the ECB is already built into the DAS.

Most of the federally funded datasets which have been “officially” released have been adjudicated, passing stringent error checks and guidelines as part of the preparation of mandated reports or tables. The institutional-level IPEDS datasets are mandated and may have imputed values for missing records or variables. None of the sample surveys include imputed data. The use of sample studies, in conjunction with IPEDS, allows for national estimates. Non-adjudicated data are not appropriate for national estimates, but are still useful for peer analysis and other type of internal studies. Likewise, the early release data which are available from the IPEDS Peer Analysis System are also appropriate for this level of study – with a notation that the data are not adjudicated or official. A researcher must know when the use of data is appropriate.

Table 7 lists Websites where users may obtain free access to federal datasets about post-secondary education that are collected by the U.S. Department of Education, the National Science Foundation, and the U.S. Department of Agriculture.

Table 7: Online Access to Federal Datasets

Agency	Collection/Tool	Website
IAED for NCES	HEGIS datasets	http://www.icpsr.umich.edu/IAED/SERIES/hegis.html
IAED for NCES	Older IPEDS datasets	http://www.icpsr.umich.edu/IAED/SERIES/ipeds.html
NCES	Older IPEDS datasets	http://nces.ed.gov/ipeds/data.html
NCES	IPEDS Peer Analysis System	http://nces.ed.gov/ipedspas/
NCES	IPEDS COOL	http://nces.ed.gov/ipeds/cool/
NCES	IPEDS Electronic Code Books	http://nces.ed.gov/ipeds/ElectronicCodebook/i2000/
NCES	Academic Library Peer Comparison Tool	http://nces.ed.gov/surveys/libraries/academicpeer/
NCES	PEQIS	http://nces.ed.gov/pubsearch/getpubcats.asp?sid=016
NCES	NEDRC Table Library	http://nces.ed.gov/surveys/npsas/table_library/
NCES	NCES Quick Tables and Figures	http://nces.ed.gov/quicktables/
NCES	Data Analysis Systems (DAS)	http://nces.ed.gov/das/hm/surveys.html
NCES	DAS Electronic Code Books	http://nces.ed.gov/das/hm/das/ecb.html
IAED for NCES	IAED SDA/DAS for RCG survey	http://www.icpsr.umich.edu/IAED/das.html#iaed
IAED for NCES	HS&B survey data through 1986	http://www.icpsr.umich.edu/IAED/SERIES/hsb.html
IAED for NCES	RCG survey data before 1991	http://www.icpsr.umich.edu/IAED/SERIES/rcg.html
IAED for NCES	NLS:72 survey data	http://www.icpsr.umich.edu/IAED/SERIES/nls.html
OPE	OPE Campus Security Statistics Website	http://ope.ed.gov/SECURITY/OPEHome.asp
NSF	WebCASPAR	http://caspar.nsf.gov/
NSF	Public Use Data Files – GSS	http://www.nsf.gov/sbe/srs/gss99pub/start.htm
NSF	Academic Institutional Profiles	http://www.nsf.gov/sbe/srs/profiles/start.htm
NSF	SESTAT	http://srsstats.sbe.nsf.gov/
NSF	Science and Engineering Indicators	http://www.nsf.gov/sbe/srs/seind/start.htm
USDA	FAEIS	http://faeis.tamu.edu/main.htm

U.S. Department of Education

While many IPEDS datasets are available for download from the NCES Website with different tools, the International Archive of Education Data (IAED) is funded by NCES to be the source for older IPEDS, HEGIS, and sample survey data. The Archive is housed in and operated by the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan. Its purpose is to “preserve all of the NCES public-use research data holdings and make these holdings, as resources permit, suitably available for research throughout the nation and the world.” Machine-readable

codebooks, documentation, and datasets are provided for free with online registration. The Archive is available at: <http://www.icpsr.umich.edu/IAED/>.

Some IPEDS datasets are available in zipped format, dating back to the early days of Gopher and the World Wide Web. Previously, the data were available on magnetic tape and on floppy disk from the National Education Data Resource Center (NEDRC), under contract to NCES. Since 2000, NCES has published all newly released data, as well as a growing amount of historical data, in its IPEDS Peer Analysis System. Currently, data from 1985 and 1990 to present are available. With the Peer Analysis System, users may select data, create calculated variables, and create unique datasets for different institutions. Another option is to download each table (or section of survey data) directly in comma-separated, CSV text format. The IPEDS Peer Analysis System is currently being redesigned and will include a new “dataset-cutting tool” for creating unique datasets. A set of tutorials about using the Peer Tool are available at: <http://nces.ed.gov/ipeds/tutorials/>.

A more user-friendly version of these data at the individual institution level, designed to meet the mandates of Congress and its College Cost Report for three years of price and cost data, is the IPEDS College Opportunities On-Line tool (COOL). IPEDS COOL also provides a link to access to Campus Security/Crime data for each institution. The Department of Education’s Office of Postsecondary Education maintains a Website with a searchable database of Campus Security/Crime data. Similarly, NCES makes data on academic libraries available with another searchable online database.

NCES sample data are available in public use files as part of Data Analysis Systems (DAS) and in restricted access files, which require a site license and special conditions of use designed to protect confidentiality. The DAS works two ways: (1) locally on the user’s computer using software and data installed via download or CD-Rom; and (2) by uploading/FTPing a DAS query from the same software to a special DAS server, where it is run and then made available for download on the DAS FTP site. While everything else is identical, the downloadable version does not allow users to create locally on the user’s computer. New versions of the DAS are released periodically as different software packages. While also provided on CD-Rom, the software available for downloading is kept more current with important updates, corrections, and recodes. The DAS for PEQIS requires a restricted use site license. In addition, IAED provides DAS-like software with its Survey Documentation & Analysis (SDA) System, but only for the Recent College Graduates, 1991 sample survey.

Information about the sample surveys has also been compiled into library of tables and is available as part of the NEDRC Table Library and the NCES Quick Tables and Figures tool. The results of NCES PEQIS surveys are also arrayed in this format. This Website application is available at: <http://nces.ed.gov/quicktables/>.

National Science Foundation

NSF's three SESTAT sample surveys (SDR, NSCG, and NSCRG) are provided to the public through a special online tool, with extensive documentation of value labels, frequencies, and changes in the data element dictionaries over time. A restricted use site license is also available for any of these sample survey datasets, plus the SED and the Survey of Public Attitudes.

Two of the three NSF institutional surveys (GSS and Academic R&D), the data collected from federal agencies, and the SED are made available through the online WebCASPAR tool. Developed originally in the early 1990's as a CD-ROM subscription service and software for analyzing datasets, WebCASPAR has evolved to a dynamic and powerful dataset tool. Just as with the IPEDS Peer Tool, users may select standard reports, create customized reports, select institutions based on criteria, and save customized reports that they create. Users who wish to create a dataset should locate the variables of interest and save the data for all available institutions.

Where IPEDS uses a list of checkboxes across multiple pages of selections to choose options and variables, WebCASPAR also allows users to visually diagram a cross-tab report, cutting and pasting different fields into the report structure to meet their needs. The GSS is available as public use datasets as well and there is a "Guide to the Data Files" that documents the data availability, distribution, and code structure.

WebCASPAR includes some important non-NSF sponsored data as well, including most of the long-standing, historical IPEDS datasets and two years of information from the National Research Council's data collection about Research-Doctorate Programs. A tutorial and discussion group are also provided.

Another useful NSF Website is the online documentation of "Academic Institutional Profiles." These profiles incorporate all data available through different institutional surveys, and array the results over time for a single institution. This is a valuable way of quickly viewing the types and years of NSF data are available for a university, in order to know whether to further explore a dataset. Another way is to view the "Data Map" feature in WebCASPAR, which documents data sources by subject, organization, variable, category, and academic institution.

Information about the Survey of Public Attitudes is available as table results in the Science and Engineering Indicators publication series and Website. A restricted use site license is available, and a public use version should soon be available on CD-Rom.

U.S. Department of Agriculture

The Food and Agricultural Education Information System (FAEIS) provides online data in two broad disciplinary clusters: agriculture, renewable natural resources and forestry; and family and consumer sciences. The data are analyzed and available in different levels of aggregation, including: national, institution, degree level, academic area of specialization, race/ethnicity, gender, region, type of institution, 1862 and 1890 Land Grant status, non-Land Grants, and by institutional membership in professional associations.

The data are provided online in different formats, including graphical images and Adobe PDF files. Users may request electronic versions of the data, at no charge, suitable for additional analyses, by emailing FAEIS staff. For more information, see <http://faeis.tamu.edu/mi~~~7.htm>.

Other National Datasets

Some survey developers make their datasets available to institutional researchers and policy analysts, although it is sometimes necessary to hand-enter information from print publications or PDF files. Of the four publishers originally involved in the Common Dataset Initiative, the College Board is the only one which will sell the data directly to institutions and states for internal use. The cost is contingent on the usage. In the past, Wintergreen/Orchard House data were sold at a much greater cost to commercial clients such as insurance companies and other publishers.

Peterson's and *U.S. News & World Report* make much of the data used in their print publications available on the Web. Users who wish to use these data should seriously consider cutting and pasting data variables by institution from these sites. Wintergreen/Orchard House supplies data on a much larger, commercial scale to interested parties, such as insurance companies and directories.

The *Chronicle of Higher Education* is an invaluable source for institutional datasets with its "Facts and Figures" and Almanac sections. For example, the *Chronicle* contains a searchable database of historical data about Crime on College Campuses, broken down by state and institution. Users may save these online pages, and readily view them in Excel, for further manipulation and analysis. Some of the other *Chronicle* data of interest include information about faculty salaries, compensation of college presidents, endowments, fundraising, research library holdings, government grants, gender equity in athletics, graduation rates, stipends and benefits for graduate students, and tuition and fees.

Institutions which participate in special data collections such as AAUP, APPA, CUPA, and disciplinary surveys usually benefit from receiving a copy of the data in print or electronic format. The AAUP Faculty Compensation Survey Report and the March/April issue of *Academe* are examples. Even if the entire dataset is not for sale, schools may obtain copies of a survey submission from peer institutions and use the results for internal analysis. Some private associations among institutions, such as the *Postsecondary Education Data Sharing Consortium*, collect data from multiple sources and repackage them for its members. For more information, see <http://hedsftp.fandm.edu/>.

Other Considerations in Manipulating Datasets

With WebCASPAR and the IPEDS Peer Tool, there is no need for the user to worry about merging multiple datasets or a single dataset across multiple years, since the software builds in this capability. SESTAT requires users to select the year of data; the system then handles merges between the SDR, NSCG, and NSCRG seamlessly, since these

are essentially the same survey with three different populations and only slight variations in content.

The manipulation of individual IPEDS, HEGIS, and other datasets is not always straightforward. Usually, the data are stored in tables, so that columns of data are related to specific questions or cells on the source survey. The variables, or fields, are named with a standard naming convention that makes intuitive sense to users, given the location of the variable on the survey instrument and/or the type of data. It is important for a researcher to know how a survey instrument has changed over time, including the availability of key variables of interest, as well as how these variables are coded, recoded, and reported in the dataset.

Where the complexity of the survey or its length dictates, multiple data files or tables may be used. Each file provides a specific section of the survey (*i.e.*, Part B – Expenditures of the IPEDS Finance Survey data file). For all IPEDS files, the Institutional Characteristics survey provides all identifying information for an institution, such as Carnegie classification, location, and control. All other IPEDS files are linked to the Institutional Characteristics, or IC, file by the institution's unique identifier or UNITID. For many NSF datasets in WebCASPAR, and for older HEGIS files, the unique identifier is Federal Interagency Committee on Education (FICE) code. For the NSF GSS survey, there are multiple submissions from an institution for each graduate program by degree level. The IPEDS Fall Staff Survey is more difficult than other IPEDS files to use as a dataset, a result of the use of multiple records for each "line."

The issue of identifying institutions is especially important when using data from multiple sources, such as IPEDS, AAUP, and a disciplinary association. Each survey developer may collect data from a different administrative unit on a campus. While UNITID or FICE may be the lowest level of identification for an institution, some departments or programs may exist in a consortia of institutions. Datasets which rely on FICE codes alone, or, worse, institutional name, may limit the user in their utility for merging to other sources. Hospitals may have their own identifiers. Systems, branch campuses, and off-site locations may be treated differently, depending upon the policy and training of the office which completed the survey at the time.

NCES and NSF both expend a great deal of effort tracking institutional changes over time, resulting from mergers, closures, changes in institutional mission, and name changes. This is the benefit of allowing NCES and NSF to keep track of multiple datasets through the IPEDS Peer Tool and WebCASPAR.

IV. Everything a researcher needs to know About a Dataset

There are a number of things about a dataset which the researcher learns only after years of use. For example, data for a peer institution may change unexpectedly across several years, for no known reason. The user may never know that the peer institution changed student information systems or that a report was prepared by different people over time using different programming techniques, who did not anticipate how the data would be

used. Another frequent problem in systems is a change in the use of value labels for a variable. An institutional researcher may request an extract to submit for a collection, never knowing that the screen design for the system may change the use of a value label, changing the meaning to something entirely different than expected.

As part of the NPEC ANSWERS tools, these relatively unknown aspects of each national dataset are better documented. Eventually, as users work with a particular dataset, they will want to make sure they understand these issues and concerns.

Table 8: ANSWERS Survey Developer Information

Everything you need to know about a dataset
<p><i>Data Collection</i></p> <ul style="list-style-type: none"> • What is the purpose of this survey? • What topical domains are covered (i.e. admissions, enrollment, faculty)? • Does the survey reference data gathered by another organization? If so, what data elements are duplication. • What specific times of the year does the survey reference? • How long has the survey been administered? • How frequently is the survey administered? • Is there an electronic data collection instrument? <p><i>Respondent Information</i></p> <ul style="list-style-type: none"> • What types of schools are surveyed (i.e. Carnegie, control, degree programs)? • What types of schools respond to the survey? • Does the survey instrument vary by type of school? If so, how • What is the average, overall response rate (over 3 years)? • What is the most recent response rate? • Which office usually receives the survey? • Which office usually coordinates the collection of survey data? • When is the survey sent to respondents? • When is the survey due? If the date is usually extended, to what date? • How long does it take a typical respondent to complete the survey (in hours)? • Are there any scheduled follow-ups to non-respondents? • Are there any enticements offered for responding (i.e. cash, free report)? <p><i>Reporting/Data Availability</i></p> <ul style="list-style-type: none"> • Do you involve schools in editing or reviewing the data? • Do the data require a significant amount of editing/cleaning? • How are missing/non-response data handled? Imputation method? • How long after the due date are the data available in some format? • Are the electronic data publicly or commercially available? If so, where, when, in what format, and at what cost? • Are reports about the data available online? If so, where, when, in what format, and at what cost? • Are reports about the data available in print? If so, where, when, and at what cost? • What are current titles of publications based on the survey data? • What is the future of the survey? • How successful is the survey in meeting the data needs of its target audience?

V. Emerging Trends in Data Collection

The past few years have seen a dramatic change in data collection, moving instruments from paper to the Web. This chapter has highlighted the major survey developers and their datasets. In preparing for future use of national datasets, however, users need to be aware of a number of emerging trends that will impact their utility.

One of the first considerations is the ongoing availability of data. Due to a major budget shortfall in agency appropriations, many IPEDS data collections were sharply cut back in

the 2000 reporting year. This cutback came at a time when NCES was moving these collections to the Web for the first time, after an exhaustive IPEDS redesign process. Therefore, users should expect that many data variables of interest may not be available for 2000. In restoring these data for 2001, NCES implemented the redesign recommendations of its NPEC Working Groups and task forces. Not all of the data cut in 2000 were fully restored. Users must check the continuity of data over time from 1999 through 2001, before expecting to find critical trend data.

Another issue impacting the availability of IPEDS data, however disseminated, involves changes in financial reporting mandated by the Financial Accounting Standards Board (FASB) which effect private not-for-private and private for-profit institutions, and the Government Accounting Standards Board (GASB), which effect public institutions. With the implementation of FASB in the late 1990s, and the imminent implementation of GASB, the comparison of public and private IPEDS Finance data is no longer possible. Also, the consistency of implementation within sectors is also in question. Therefore, some financial data are not yet available for this time period. Ongoing financial data may, or may not, be collected, depending upon the budget problem of 2000, the move to Web collection, the implementation of FASB/GASB, and the implementation of the IPEDS redesign.

Several new data collections are becoming available or are in the pilot phase. The IPEDS Employee by Assigned Position Survey was voluntary in 2001; however it is required in the 2002-03 cycle. A survey of Instructional Activity, allowing for measures of productivity, is being piloted as part of an NCES Working Group. It is possible that these data could become part of IPEDS in the future. Some institutional characteristics data, such as average test scores, which have been part of the college admission guide's collection, are becoming part of IPEDS.

Posing a greater challenge to data comparability over time, the Office of Management and Budget (OMB) will require institutions to begin reporting about students and faculty using new race and ethnicity coding. The NSF GSS already incorporates this change. However, it is not yet completely clear that the final version of the coding will be what the GSS implemented. All trend and historical data about race and ethnicity will be lost once these new reporting requirements are fully implemented. During those years in which institutions struggle to adequately report the data until their systems are in place, the results may not be comparable.

Keeping track of changes

In order to keep track of these changes, dataset users should watch closely the NCES, NSF, and AIR Websites. As part of its AIR/NCES/NSF grant program, entitled "Improving Institutional Research in Postsecondary Educational Institutions," the Association for Institutional Research (AIR) promotes "opportunities for postsecondary education professionals and doctoral students to conduct research utilizing the national databases" through grants, institutes, and a post-masters certificate program. Additional information is available at: <http://airweb.org/awards.html>.

One key resource for using the national datasets is the annual NCES/NSF Summer Database Institute, which trains approximately 40 fellows per year with a “combination of instruction on the content and uses of the NSF and NCES national data sets relevant to postsecondary education, and policy seminars focused on national postsecondary education issues.” For those who cannot attend the Institute, a number of workshops and presentations about NCES, NSF, and other data collections are held at the annual AIR Forum and other regional and association conferences. For federal, SHEEO, and affiliated association staffers, the annual SHEEO/NCES Network has been in place since 1976, with an annual meeting and other important activities. One of the purposes of the SHEEO/NCES Network is to “make national and state data collections valuable and relevant to policy-makers.”

The Higher Education Data Policy Committee (HEDPC) of AIR works to promote communication regarding a number of data issues which impact users. The HEDPC Website provides updates on the status of its projects. Additional information is available at the AIR Website: <http://airweb.org>.

Another important feature of the AIR Website is the Internet Resources for Institutional Research feature. Offered since 1995 and housing several thousand links in over eighty categories, this Website is the oldest and most complete source of online information about postsecondary education and is available at: <http://airweb.org/links>.

Finally, dataset uses should follow the work of the National Postsecondary Education Cooperative (NPEC). Congress authorized NCES, in 1994, to establish the Cooperative. Its mission is “to promote the quality, comparability, and utility of data for postsecondary decision-making at the national, state, and institutional levels.” Additional information about NPEC is available at: <http://nces.ed.gov/npec/>.

VI. Summary

This chapter documents the many types of postsecondary education datasets which are available to researchers, along with how they may be accessed and better understood. While the datasets and software tools are very complex, they are also very rich topically, with information to support many kinds of policy, research, and analytical studies. Much of the potential of these datasets has not been utilized.

Researchers are encouraged to make use of existing, national, sample and institutional datasets whenever appropriate, especially those from NCES and NSF. An enormous effort has been undertaken to transform the collection and dissemination of data from the printed survey form and magnetic tape of the early 1990’s to the online data entry forms and analytical tools of today. With an exciting array of new data and software, researchers are encumbered with the responsibility to consider what these datasets have to offer.

References

Association for Institutional Research. (2002). *NCES and NSF Postsecondary Datasets: Descriptive information about the purpose, respondents, key variables, target population and sample frame from popular NCES and NSF surveys*. Tallahassee: Association for Institutional Research.

Milam, John H., Jr. (1999). "Using the National Datasets for Faculty Studies." *AIR Professional File*, No. 70. Winter, 1999.

National Center for Education Statistics. (2002). *Digest of Education Statistics, 2001*. Washington, D.C.: U.S. Department of Education.
<http://nces.ed.gov/pubs2002/digest2001/index.asp>.

Russell, Alene B., and Susan B. Winter. (2001). *Compendium of National Data Sources on Higher Education*. Fifth edition. SHEEO/NCES Communication Network. Denver: State Higher Education Executive Officers.