

# Understanding the Core Data Service

Higher education continues to experience unprecedented pressure for accountability from both internal and external constituencies, from trustees to campus administration to prospective students and their parents to governmental agencies. In recent years, these accountability demands “have been especially targeted at information technology, putting strong pressures on IT leaders to explain and justify the costs and benefits of the expenses associated with their areas.”<sup>1</sup> Fundamental to such efforts is having reliable data about information technology practices, structures, and expenditures at comparable institutions for benchmarking purposes.

Finding such useful and relevant comparative data for IT units in higher education has long been a challenge, and a number of data-collection activities arose through the years to meet this need. (See Appendix A for the historical context from which the current EDUCAUSE Core Data Service arose.) Five years ago EDUCAUSE determined the need for a somewhat different approach from existing data collection efforts and thus developed and launched a program called the EDUCAUSE Core Data Service (CDS), which consists of an annual survey instrument that collects data about information technology environments and practices on (primarily) member campuses; a Web-based, interactive database service available to all who complete the survey through which they can access data contributed by their peers to help benchmark, plan, and make decisions about IT on their campuses; and an annual, publicly available summary report about campus IT environments based on data contributed through the survey.

This *EDUCAUSE Core Data Service Fiscal Year 2005 Summary Report* is the fourth report published as part of the CDS program.<sup>2</sup> Before delving into the five major sections that follow this introductory section (each of which parallels and summarizes data from a section of the core data survey), we encourage you to read on to fully understand the CDS program, especially

its underlying principles, appropriate use policies, and methodology (including use of Integrated Postsecondary Education Data System, or IPEDS,<sup>3</sup> data), and how data are analyzed and presented in this summary report.

## Underlying CDS Principles

A defining characteristic of the EDUCAUSE CDS is its collection and presentation of data *identifiable by institution* in the interactive database component of the service. The level of participation in the program is evidence that the value of being able to select a specific comparison group of similar, peer institutions outweighs any reluctance participants might have to disclose identifiable data. (See Appendix B for a list of 2005 survey participants.) The willingness of the community to share what until the inception of the CDS had been largely unavailable financial data has allowed this service to approach the status of a breakthrough application.

A second fundamental principle of the program is that only those campuses that complete and submit the survey each year are eligible to log into the interactive database site. Nonparticipating campuses do not have access, nor do corporations, researchers, agencies, associations, the media, or the general public. However, EDUCAUSE feels an obligation to provide some overall data analysis to member campuses that do not participate, as well as to the vendor community that supports the association, and thus we publish this annual summary report for distribution on a complimentary basis to the entire EDUCAUSE membership.

A third important element of the CDS is its appropriate use policy and the efforts expended to ensure that all survey participants are well informed about the conditions and terms of use of the data captured through the CDS survey. Access to the database service is not only restricted to participating campuses but further restricted to individuals on those campuses who have been authorized by their cam-

pus to use the database. Such authorized access is provided through an EDUCAUSE username and password issued to authenticated individuals recognized by the CDS system. The CDS has a strong copyright and appropriate use policy (see <[http://www.educause.edu/coredata/use\\_policy.asp](http://www.educause.edu/coredata/use_policy.asp)>) expressly to protect the information of participating institutions. Anyone authorized to access the database must “click through” and agree to all of the terms and conditions of use before gaining that access. Any campus found in violation of the terms and conditions of use will be penalized by loss of participation privileges in the CDS, and EDUCAUSE may take legal action against any party who accesses or uses database content or data without authorization.

Finally, a note about the trust relationships and partnership we enjoy with our corporate members: While use of the CDS database is restricted to campuses that complete the survey, thus de facto precluding vendor participation, some of the campus individuals who complete the survey are in fact employees or contractors with corporations that have facilities management contracts with their campuses. We contacted the companies known to offer such services (Blackwell Consulting Services, CampusWorks, SunGard Higher Education), and they graciously agreed in writing not to seek access to the service or survey data, realizing that it is strictly for campus consumption for planning and institutional analysis. Further, they agreed that if any data did come into their hands, they would not use it. It is this level of true partnership that we in higher education are fortunate to enjoy with our corporate community.

## **Methodology**

All EDUCAUSE member campuses that have an IPEDS unit ID number as well as international member institutions (which do not have such numbers) are invited to complete the core data survey through an e-mail message sent annually in January to the primary representative at each member campus. We also invite schools that are not members of EDUCAUSE to participate in the CDS if they

are members of affinity groups (such as the Council of Independent Colleges, the League for Innovation in the Community College, and others) as well as any campus that expresses an interest in completing the survey. In January 2006, more than 2,600 campuses were invited to participate in the 2005 survey.

In the case of an institution with a Carnegie classification<sup>4</sup> of “system,” individual member campuses within the system are invited to complete the survey, provided they have an IPEDS unit ID number. A multicampus system with a single unit ID is invited to complete the survey as a single institution. System or district offices (except those that have a single IPEDS unit ID) are not eligible to complete the survey; however, if 40% of the campuses within the system or district complete the survey, the system or district office becomes eligible to access the interactive database service.

Access to the survey is provided through an authorization system that gives such access initially to the individual designated as the primary or key representative in the EDUCAUSE records database at the time the invitation to participate is extended. That individual is invited to manage the completion of the survey on his or her campus or to designate another individual or individuals to do so.

All data captured by the core data survey are submitted electronically through an easy-to-use Web-based interface that enables respondents to answer the approximately 50 questions over time; that is, they can enter data, save them, and return to the site at another time to enter more data or change data already entered. Participants are given about two months to submit the survey, which can take anywhere from several hours to several days to complete, depending on the ready availability of the campus data requested. (See Appendix C for a copy of the 2005 survey.) Note that all financial data sought through the core data survey are for the previous fiscal year, so actual funding/expenditures rather than projected budgets are captured. For example, the survey launched in January 2006 sought financial data for fiscal year 2004–2005 and thus is referred to as the 2005 core data survey. Once a campus submits

its survey, data cannot be changed except by special request, for example, in the case of incorrect data having been submitted.

Embedded throughout the survey are a variety of pop-up and linked help notices, electronic navigation to a glossary of terms and definitions, and other aids to clarify questions and to obtain consistent responses. (A list of the glossary terms appears in Appendix D of this summary report.) An audit system provides red-flag messages to respondents if inconsistent data are entered, giving the respondent an opportunity to correct data after viewing an explanation of why the data appear to be problematic.

### **Use of IPEDS Data**

EDUCAUSE information systems enable automatically matching respondents with their corresponding IPEDS data, so these elements do not have to be entered by the respondent. Based on data reported by U.S. colleges and universities through IPEDS for 2004 (the most up-to-date IPEDS data available), the number of FTE faculty, number of FTE students, total student headcount, gross general institutional expenditures, and type of institutional control (public or private) are matched in the database, as is the Carnegie classification for each institution.

Despite the best of intentions, IPEDS data have proved to be inconsistent and inappropriate for much of what we intended to accomplish. The first year of the CDS, we conducted extensive analyses of our survey data and selected IPEDS data (faculty FTE, student FTE, and total institutional expenditures), both to ensure data integrity and to prepare the first core data summary report. Through those analyses, we determined that some of the campus IPEDS total expenditures data were inconsistent, and thus ratios using that data element, as well as approximations of the former educational and general (E&G) budget of an institution, could not be used due to their unreliability.

Various faculty and student ratios based on IPEDS data were included in that first summary report, but subsequent analyses (after the printing of the report) showed some serious prob-

lems with the IPEDS faculty data. Thus we advise that the two ratios presented in the 2002 summary report that were based on IPEDS faculty data should be considered questionable. These ratios are no longer included in our summary reports or in the ratio section of the online database component of the CDS. However, the actual IPEDS data are available through the demographic feature of the database for those who wish to include them in their analyses.

In researching the way financial and faculty data are reported to IPEDS, we collaborated with several commercial vendors, the Department of Education, the National Center for Higher Education Management Systems (NCHEMS), and other groups that are actively using comparable data. We learned that these problems are endemic with IPEDS and that there are no easy workarounds. Please see the item dated March 2, 2004, at <<http://www.educause.edu/coredata/news>> for details of the issues with IPEDS data on total institutional expenditures and faculty FTE numbers, as well as for suggestions for using IPEDS data, with caution, in your campus analyses.

Note that the 2005 core data survey requested two additional data points (albeit on an optional reporting basis) that were needed for institutions that have previously participated in the COSTS Project to calculate benchmarks to which they had become accustomed. The requested data were total number of headcount employees (including faculty) reported the previous year to IPEDS and total campus expenses (not including financial aid) reported the previous year to IPEDS. Those who answered the latter question were also asked to indicate which accounting standards their campus used (FASB, Financial Accounting Standards Board, or GASB, Governmental Accounting Standards Board). Collection of these self-reported IPEDS data enabled the incorporation of seven additional benchmarks into the interactive database service component of the CDS.

### **How Data Are Presented in this Summary Report**

Data for this summary report are reported by 2000 Carnegie Classification, but we have

combined like Carnegie categories for ease of reporting and for manageable data presentation in the tables. In doing this, we ensure that by combining groups we do not lose important distinctions. Appropriate statistical tests are conducted with a large number of variables in the data to determine if consistent and meaningful differences exist between like categories. Within the Carnegie categories, tests are run to determine if such categories can be combined. In all these sets of analyses, for the fourth consecutive year, no significant patterns were identified when the size differences in the schools were controlled for. This was also the case when controlling for public versus private control.

Thus, throughout this report—with very few exceptions—the data displays focus on the following combined categories: BA, which combines Baccalaureate Colleges-Liberal Arts, Baccalaureate Colleges-General, and Baccalaureate-Associate's Colleges; MA, which combines Master's Colleges and Universities I and Master's Colleges and Universities II; DR, which combines Doctoral/Research Universities-Extensive and Doctoral/Research Universities-Intensive; and AA, which includes all schools with a classification of Associate's Colleges (community colleges, technical colleges, junior colleges, and other colleges that grant associate's degrees). Definitions of these 2000 Carnegie classifications are included in Appendix E. Our category of OTHER includes Tribal Colleges and schools in the Specialized Institutions Carnegie class (such as law schools, health-related institutions, art schools, and so forth), as well as participating international institutions, which do not have Carnegie classes assigned because that is a uniquely U.S. schema.

The purpose of this report is to provide aggregate data in simple form for those who do not have access to the interactive database service. In our analyses we have not tried to provide every possible cut on the data, but rather some summary data that we believe will be useful to the public. Keep in mind that the database service component of the CDS allows for viewing data much more discretely.

The service offers filters, sorting tools, graphing tools, the ability to see trend data comparing last year's and this year's data (see details about trend analyses below), and a sixth section that provides automatically generated ratios in 14 areas.

We urge readers who have access to the database service to use the service rather than this report for benchmarking purposes for a more refined and accurate picture than the tables in this report can provide.

### **Core Data Survey Participation**

A total of 933 institutions had submitted the 2005 survey when we froze the data set early in June 2006 to do the analyses for this summary report. Submissions continued to come in throughout the late spring and summer and likely will continue for the rest of the 2006 calendar year. As of November 1, 2006, 950 campuses had submitted the 2005 survey. This level of participation is up about 2% over the 928 institutions that participated last year. (Note that last year there were 890 schools in the frozen data set.) This year there are also increases in the number of schools participating in several of the Carnegie groups, with a 1% increase in doctorate-granting institutions, a 5% increase in baccalaureate colleges, and a top increase of 9% in master's colleges and universities.

The level of participation among statewide and multicampus systems and districts also increased this year. As was the case last year, many system offices encouraged the fullest participation of their member campuses. This year nearly three dozen systems or districts achieved at least a 40% participation rate in the CDS (with several reaching or approaching 100%), as compared with 30 last year.

### **Trend Analyses**

An electronic set of tools is available to those who use the interactive database service component of the CDS so that they can see trends within specifically defined peer groups or other categories of analysis for the past two years. Using these tools, users can determine if they want to compare the data of all 2004 and

2005 survey participants or if they want to compare data of just those institutions that completed both of the surveys (that is, institutions in the matched data set). In the latter case, actual change is more confidently ascertained, whereas in the analyses that would compare all participants from each year, some of the change is likely to result from a different sample, possibly leading to false conclusions. In this summary report, the narrative attempts to highlight key trends when they are seen to be important, but only comparing data for the 749 schools that are in both this year's and last year's frozen data sets.

When comparing data for all of these 749 schools, finding statistical significance is likely to occur quite frequently because of the large sample sizes. Many of the most interesting changes do not occur across the board, however, but are patterns specific to community colleges, research institutions, or other Carnegie groups. When examining those subgroups within the matched data set, sample sizes become fairly small, and statistical significance is harder to find. In some of those cases, the narrative in this summary report will note these changes (which may or may not be due to chance), even though statistical significance was not found, simply to hypothesize a possible trend of special interest.

### **The Fallacy of Relying Only on Input Measures**

We began this introductory section by proposing that the collection of IT-related data is important to help campuses plan more effectively by virtue of having access to information about IT infrastructure, funding, and management practices of schools similar to themselves. But the problem with IT benchmarks of any kind—and the CDS is no exception—is that these input comparisons are too often used to convince decision makers to keep pace with their peers and that more is better where technology is concerned.

This effort to “keep up with the Joneses” is ultimately an inflationary pressure that can be dysfunctional, acting as a negative driver. Such pressure and focus on input measures is

a fallacy that higher education is finally beginning to recognize. Rather than engaging in an “arms race,” we need to focus on effectiveness—trying to determine which institutions seem to be doing the best job with the fewest resources, with an eye toward understanding the environment and practices that make this possible. Hawkins and Barone made the case for a new kind of assessment model that not only uses input measures but also recognizes the even greater importance of evaluating outcomes in higher education:

Although...efforts [using input measures] may have leveraged additional funds (appropriately or not), they do not include measures that offer insight into how technology is enabling new and better research, whether or how technology is enhancing teaching and learning, or whether administrative functions are easier for students to access or less expensive to operate. The problem is that in order to effectively measure the success and/or value of an IT investment, we must come to grips with evaluating these functional outcomes of the college or university. However, we have thus far successfully avoided grappling with these difficult challenges of assessing learning outcomes, administrative efficiency, effectiveness, and so on. Without working in tandem with others on campus to identify and evaluate these outcomes and then to understand and describe the enabling role of IT in facilitating these accomplishments (or the failure thereof), we will never be able to reasonably and meaningfully assess the return on IT investment.<sup>5</sup>

Some might suggest that the EDUCAUSE CDS may contribute to the fallacy of overvaluing input measures, but we would counter such an allegation on several fronts:

- First, this kind of application is in very high demand by our members for a host of reasons, among them being

able to understand where the market really is and what other campuses are actually doing, in order potentially to reduce the pressures on growth and expansion.

- Second, even if legitimate outcome measures were available, we would still require input measures to understand the effectiveness equation. Efforts such as the CDS are necessary but not sufficient to achieve the ultimate goal of defining standards of optimal achievement of goals.
- Third, the CDS database service has the potential to dispel the myths surrounding IT funding and investment by presenting detailed data that present a more accurate and reliable picture of campus IT environments.
- Fourth, the interactive service is providing a useful network to help participants find and communicate with colleagues like themselves, who have similar systems and characteristics and who are facing similar challenges, and to learn from them.
- Fifth, the CDS has the potential to promote more congruity in campus IT funding models, provide models for IT organization and support, identify exemplary processes for allocating and expending resources (both human and financial), and promote more effective IT management overall through prompting more widespread tracking of IT expenditures (whether these occur internally or externally to the central IT unit) at higher education institutions.

We believe that the CDS also has the potential to create a different sociometry for the IT community, replacing the casual inquiry to a listserv for information with a more informed method of obtaining comparative data. All too frequently a concerned member will post a query on the CIO listserv asking, for example, Who out there has or is considering having the library report to the CIO? A few folks respond, but the results are serendipitous and incom-

plete, based on who happens to be reading the listserv at the time, whether or not the respondents are from similar types of institutions, and so forth. That is but one question the CDS can answer, filtering responses based on criteria such as Carnegie class, FTE enrollment, public versus private control, and even institutional budget, until a short list of the most appropriate schools for comparison appears. Clicking on any school on the list will link to the EDUCAUSE member directory, where all of the representatives to EDUCAUSE for that campus are listed, including contact information. This facilitation of communication between and among members of the community, based on information about areas of common interest or challenges, has from the beginning been a key objective of the EDUCAUSE CDS.

As illustrated by the excerpt from Hawkins and Barone, there is a clear and pressing need for higher education to focus on outcome goals, and EDUCAUSE has both been advocating in this arena and partnering with other higher education organizations to advance this agenda. We fully recognize that our core data program is not the endgame, but it is an important part of the total picture. It is our hope that eventually our service will be part of the analysis in determining the most efficient methods and effective practices for achieving important output objectives and goals.

## Notes

1. B. L. Hawkins and C. A. Barone, "Assessing Information Technology: Changing the Conceptual Framework," in *Organizing and Managing Information Resources on Your Campus*, P. A. McClure, ed. (San Francisco: Jossey-Bass, 2003), pp. 129–145.
2. The 2002, 2003, 2004, and 2005 CDS summary reports are available for free download in PDF form on the EDUCAUSE Web site at <<http://www.educause.edu/core/data/reports>>. Print copies of this 2005 report are available for \$10 each as long as the supply lasts.
3. The Integrated Postsecondary Education Data System (IPEDS) is a single, comprehensive, data-collection program designed to capture data for the National Center for Education Statistics (NCES) for all institutions and educational organizations whose primary purpose is to provide postsecondary education in the United States.

IPEDS collects institution-level data in such areas as enrollments, program completions, faculty, staff, and finances. IPEDS data reporting requires the extensive effort of a variety of offices on any campus, and this is the “official” information the college or university stands behind, used by the federal government.

4. In 1970, the Carnegie Commission on Higher Education developed a classification of colleges and universities to support its program of research and policy analysis. Derived from empirical data on colleges and universities, the “Carnegie Classification” was published for use by

other researchers in 1973, and subsequently updated in 1976, 1987, 1994, 2000, and most recently in 2005. With the 2005 revision, the single classification system was replaced by a set of multiple, parallel classifications. The original classification framework—now called the basic classification—has also been substantially revised. For details about those revisions, see <<http://www.carnegiefoundation.org/classifications/index.asp>>. This CDS summary report uses the basic classification system from 2000, for the sake of simplicity.

5. Hawkins and Barone, *op. cit.*, p. 133.