



May 2004

California Postsecondary Education Commission

# Information Collection and Dissemination Program

*Enhancing access to education data*

*The California Postsecondary Education Commission is a citizen board established in 1974 to coordinate the efforts of California's colleges and universities and to provide independent, non-partisan analysis and recommendations to the Governor and Legislature.*

*More information on the Commission, including links to Commission publications, is available at [www.cpec.ca.gov](http://www.cpec.ca.gov). Direct access to the Commission's online data system is available at [www.cpec.ca.gov/OnLineData/OnLineData.asp](http://www.cpec.ca.gov/OnLineData/OnLineData.asp).*

## Contents

1. Background .....	2
2. First Focus: Students.....	3
3. Enhancements to Student Data.....	4
4. Other Data Collected .....	5
5. Data Exchange and Collaboration .....	6
6. Issues facing the program .....	7
7. Recommendations.....	8
8. Bibliography.....	11

*Do not seek for information of which you cannot make use.*

*– Anna C. Brackett*

The California Postsecondary Education Commission (CPEC) has collected and provided information about education – particularly higher education – in California since 1976. As a “one-stop-shop” for information about education beyond high school, the CPEC Information Collection and Dissemination Program is well known for its commitment to data quality, responsiveness to users, and the consistency and structure of the data.

The increased demand for timely, accurate, and accessible data to inform policy decisions on demographic changes, accountability initiatives, and on-going fiscal constraints in California places a premium on information. Education stakeholders increasingly use CPEC as the resource for comprehensive and objective information.

Despite its many strengths, the Commission's database system could be made even more valuable to education stakeholders if the recommendations contained in this report are implemented.

## 1. Background

The Commission's Information Collection and Dissemination Program has its origin in the 1974 enabling legislation calling for the development and maintenance of a "postsecondary education information clearinghouse." After consultation with all interested parties, the Commission established its Information Systems Program in 1976.

CPEC's Information Collection and Dissemination Program was based on the following principles:

1. *Responsiveness to users:* The program was designed to respond to most requests for information within five working days.
2. *Limited scope:* In order to minimize the reporting burden, the Commission agreed that the segments would only provide information that met all three of these criteria:
  - There was documented need for the information;
  - The information requested could be used as a "reliable" indicator; and
  - The information requested could be used to provide meaningful comparisons among segments, programs, or activities.
3. *Centralized data collection:* Generally, the Commission would obtain the information it needed from the public segments' systemwide offices and not the campuses. Since the Independent segment has no systemwide office, the Commission would use data collected through the National Center for Education Statistics (NCES) survey system in order to minimize the reporting burden on these institutions
4. *Systematic feedback to information providers:* The Commission agreed to audit all data submitted and confirm with the systemwide offices its accuracy before using it in Commission studies and publications.
5. *The database system would be based on collecting and providing information for educational policy analysis rather than managing an educational system:* Typically, the Commission would collect information that reflected activity in postsecondary education at defined points in time. This information would be used to inform policy discussions and was not intended to include the details required to manage student flow, faculty assignments, and the like. (See sidebar: *A Comparison between a Data Analysis System and a Management Information System* on page 3)
6. *Commitment to collect defined information:* The Commission's initial effort would be centered on establishing (1) a comprehensive fall-term enrollment file; (2) a full-year degrees conferred file; (3) an inventory of degree/certificate programs; and (4) an institutional directory.
7. *The database is designed to be flexible and produce reports and compilations in response to requests from a wide variety of users.*

**A Comparison between a Data Analysis System and a Management Information System**

The design of a database influences its usefulness to the users. It is important that the characteristics of the database match the expectations of the users. The Commission uses a database model that allows for effective and efficient data analyses as opposed to a database that provides for the day-to-day management of an education system. A better understanding of the capability of each type of database is obtained by comparing the characteristics of the two models:

**Data Analysis System**

1. Used by "knowledge workers" for analysis
2. Snapshots of data at time intervals
3. Data access is *ad hoc*
4. Data is filtered, summarized, condensed
5. Data is from multiple sources

**Typical questions answered by a Data Analysis System:**

1. What are student enrollment trends for the next ten years?
2. How many students earned a master degree in 1994?
3. Rank the College-going counts by Region.

**Management Information System**

1. Used for day-to-day management
2. Data always current
3. Data accessed continuously
4. Single version of data

**Typical questions answered by an MIS System:**

1. What are the test scores from yesterday's quiz?
2. Who is out sick today?
3. In which room is Biology 101 taught?

**2. First Focus: Students**

The Commission collects information on students from two types of data sources: (1) an individual record for each student enrolled, and (2) data aggregated by common student characteristics. Individual records allow the analyst to summarize the data in a variety of ways not possible with aggregated data. For example, using individual records allows the analyst to create any groupings of the data for analysis while aggregated data forces the analyst to use pre-defined groupings.

Public colleges and universities submit an individual data record for every student enrolled and for every degree awarded. This data format allows the Commission the greatest amount of flexibility in producing reports.

Data about students attending non-public colleges and universities are obtained through the Federal Department of Educa-

<b>Information collected about students</b>	
<b>Citizenship Status</b>	<b>High School of Origin</b>
<b>Credit Load</b>	<b>Institution of Attendance</b>
<b>Degree Obtained</b>	<b>Permanent Residence</b>
<b>Enrollment Status</b>	<b>Student Level</b>
<b>Ethnic Status</b>	<b>Student Major</b>
<b>Fee Status</b>	<b>Student Program</b>
<b>Full/Part-Time Status</b>	<b>Student's Year of Birth</b>
<b>Gender</b>	<b>Transfer Institution</b>

tion NCES survey system. The data gathered from this survey system is aggregated and, thus, is more limited in its flexibility and usability.

Data on first-time freshmen and community college transfer students is also gathered from degree-granting, non-public colleges and universities (Independents) through a CPEC on-line survey system.

### **3. Enhancements to Student Data**

As data users – both internal and external – have become more demanding, the number of data elements on students has expanded. Recent additions include information about a student's disability status, year of high school graduation, student age, and expanded racial/ethnic categories. All of these items were added so that policy makers could have a better understanding of the effects of their decisions on the students of California.

Another enhancement to the Commission's student data collection effort is the addition of a student identifier. Assembly Bill 1570 (Chapter 916, 1999) authorized CPEC to augment its existing student information database with a unique student identifier for each student record.

As a first step towards building a student information system with longitudinal analysis capability, the University of California (UC) and the California State University (CSU) submitted a limited set of data on students they identified as California Community College (CCC) transfer students beginning with students who transferred in Fall 2000. The Community College Chancellor's Office has provided access to their student data so the Commission can begin to build this enhanced system.

The addition of a student identifier enables the Commission to study the movement of individual students into and through the public segments of higher education. The Commission can better understand and report on time to degree, dropout and stop-out patterns, transfer patterns, and concurrent enrollment. The information reported from the enhanced system provides information necessary for policy makers to determine how well public colleges and universities are responding to state policy priorities and accountability goals.

Even with the addition of the student identifier, there continue to be gaps in the information collected that would allow for a better understanding of student success or failure. Some of these gaps include information about Socio-Economic Status of the student and his/her family, course taking patterns, and student outcomes other than the degree obtained.

Given the concern raised by the Department of Finance about the cost of adding the student identifier and the concern raised by the segments about an increased reporting burden, any proposal by the Commission to increase the number of data elements collected will have to address these issues.

In the interest of building a seamless information system about student flow and student outcomes, the Commission continues to work with its colleagues in K-12 education to ensure compatibility between the CPEC information system and the California Student Information System (CSIS).

## 4. Other Data Collected

While the initial focus of the information collection program was on students, the Commission has added other data useful in understanding issues impacting education in California. The Commission has been the designated agency in California to coordinate federal data collection. The Integrated Postsecondary Education System (IPEDS) provides other data in the CPEC database. Some of this data is identified below.

### **Institutional Characteristics**

The Commission maintains an inventory of California postsecondary degree-granting Colleges and Universities including off-campus centers. This database includes over 100 data elements describing each institution. These data are used to produce the *Institution Profiles* ([www.cpec.ca.gov/CollegeGuide/Profiles.asp](http://www.cpec.ca.gov/CollegeGuide/Profiles.asp)), *Guide to California Colleges and Universities* ([www.cacollegeguide.com](http://www.cacollegeguide.com)), and are used in *County Profiles*, *Legislative Profiles* and virtually all of the other reports produced by the Commission.

### **Faculty and Staff**

The Commission collects data on faculty and staff including their classification, gender, ethnicity, contract length, and tenure status. These data are gathered through the NCES survey system and are generally available for all public institutions and most of the degree-granting Independent Institutions. Currently, access to these data is by special request although there are plans to provide direct access on the CPEC website.

### **K-12 Data**

Preparation for college begins when a child first goes to school. Understanding the factors that influence whether students are prepared for college attendance is critical to effective policy analysis and recommendations. The California Department of Education (CDE) is the source for much of these data. Among the data included in the Commission's database are:

- Public K-12 Enrollment
- Public High School Graduates
- College Preparation Course (a-g) Completions for Public High Schools
- A complete crosswalk of all codes used for schools by various agencies
- Private School Data (Less complete than public school data)

CPEC continues to explore links between K-12 data and higher education data.

### **Workforce Data**

There is growing interest in outcomes after students obtain a degree. There is also widespread recognition that education at all levels should be better aligned with workforce needs. The United States Department of Labor has shown that today's graduate will make three major career changes before retirement. Data also shows that a percentage of future jobs will require some training beyond high school. All of these factors mandate that the Commission gather data that documents trends and patterns in the labor market. Information about labor market trends, employment patterns, unemployment rates, and the like are collected from the California Employment Development Department, the California Department of Finance, and the U. S.

Department of Labor. These data are used in the Commission's *Legislative Profiles* and *County Profiles*. Currently, access to these data is by special request although there are plans to provide direct access on the CPEC website.

### **Census Data**

The decennial United State Census provides a wealth of information about the characteristics of the population including Socio-Economic Status (SES). The Commission has added this data to its information collection so that it can provide a basic description of the population of the State. Census data has been used in various Commission reports.

### **College Preparation Data**

In conjunction with the data collected from the California Department of Education, the Commission also gathers data on the various Outreach and College Preparation Programs offered throughout the State. These data can be accessed directly from the CPEC website at <http://www.cpec.ca.gov/OnLineData/OutreachSearch.asp>.

### **Prison Data**

The Department of Corrections provides the Commission with data on the prisoner population by gender, ethnicity, and county. This information is used in comparison with college-going rates as a measure of the spectrum of success and failure of California's education system.

### **Miscellaneous/Supporting Data**

The Commission receives information from a variety of other sources to supplement and enhance its various databases. Among the data collected are U. S. Postal Zip Codes, California geographic information, geocode information, Legislative District information, and information from various other Federal, State, and local sources.

## **5. Data Exchange and Collaboration**

A component of the Commission's Information Collection and Dissemination Program that has grown in importance is the integration of information from a variety of sources. This integration of data has not only enriched the reports and analyses produced by the Commission but it also allows the sharing of information among various agencies. This is very much in keeping with the Commission's view that its database system is a resource to be shared by all those interested in education in California.

A good example of this data exchange and collaboration is the Commission's crosswalk system, which maps coding structures and definitions among various agencies. Among the items that can be cross-walked are school/institution codes, ethnicity of students and faculty, gender, regions, and degree types. In order to facilitate collaboration, CPEC staff participate in a number of intersegmental data exchange groups.

## 6. Issues facing the program

Despite the Commission's successes in building its Information Collection and Dissemination program, it still faces a variety of challenges:

### Loss of staff

In 2002, the Commission's Information Systems Unit had fourteen staff. Because of reductions to the Commission's budget, staff shrank to four and one-half positions by 2004. While the impact of these losses has been substantial, they have been mitigated to some extent by leveraging technology. (See sidebar: *Leveraging Technology: Doing More with Less* below). However, it must be recognized that technology can only be leveraged so far and there is no substitute for staff when judgment, creativity, and imagination are required.

### Quality Control

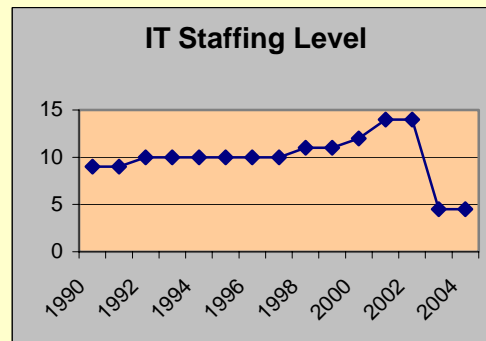
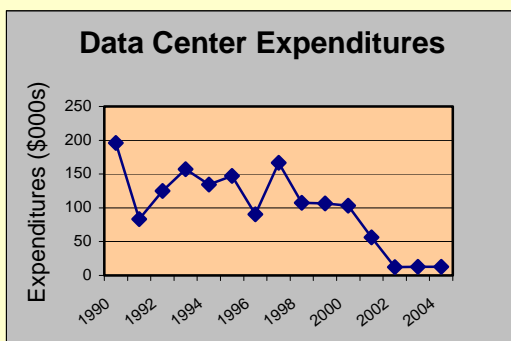
One of the personnel positions lost was dedicated to maintaining the quality of the data. When this position was filled, the Commission was proactive in maintaining data quality. Even with reduced resources, CPEC remains committed to providing quality data.

### Hardware/Software

The degree to which staff depends on technology to maintain its capacity to achieve CPEC's myriad responsibilities requires that its hardware and software be current, supportable, and reliable. Some of the Commission's data hardware are becoming old and will require replacement soon. In addition, some of the software is no longer supported by the manufacturer and must be replaced or upgraded.

#### Leveraging Technology: Doing More with Less

Throughout its history, the Commission has used available technology to build and maintain its Information Collection and Dissemination program. As the charts below indicate, the Commission's Information and Dissemination Program provides information and data today using fewer resources and at less cost. Response time for most data requests have been reduced from 5 – 7 days to a matter of minutes (via the Commission's website). Finally, the availability of the data is now "24/7" versus the business week model of 8AM – 5PM on weekdays.



### **Compatibility**

Postsecondary institutions do not use a single standard Management Information System (MIS) even within the same system. Definitions of data elements vary between the systems. Local priorities may override state level concerns and reporting requirements. At any given point in time, some institutions are upgrading or changing their software or hardware. All of these can affect the timeliness and accuracy of the data reported.

K-12 institutions use a paper survey instrument to report data. This information is, at best, a snapshot in time on a particular day.

### **Segmental Responsiveness and Data Availability**

Responsiveness to Commission requests for new types of data is subject to its perceived use by the Commission and how it might negatively impact the provider of the data. Independent Institutions provide limited sets of pre-compiled data, which have limited usefulness. Generally, the "State Approved and Exempt" subset of private institutions do not submit any data. This is not expected to change any time soon. Collecting this data is important in order to have a complete picture of education beyond high school in California.

Commission staff makes every attempt to keep users of the data informed about specific source problems through its Caveats Page (<http://www.cpec.ca.gov/OnlineData/Caveats.asp>).

## **7. Recommendations**

In order to sustain and build upon the Commission's Information Collection and Dissemination Program, the following recommendations are offered:

### **Staffing**

The number of staff needed to maintain the Commission's Information Collection and Dissemination Program is dependent on a number of factors. These include the expertise of staff, the depth and breadth of the Program, user demands, available technology and training. CPEC Information Systems staff consists of four and one-half positions with over seventy years of experience. This level of experience in conjunction with leveraged technology has permitted the Commission to maintain the existing Information Program with only occasional delays. While user demand has increased, the Commission's website has mitigated the impact on staff workload.

Other CPEC staff have had to become more self-reliant because IT Help Desk activities have ceased. On the continuum of a Reactive to Proactive scale, a reduction in InfoSystems staff forces a slide toward the Reactive end of the scale. Focus is maintained on Students while other data projects are given a lower priority status.

At a minimum, three additional positions should be added to support the Commission's Information Collection and Dissemination Program. Two of these positions would be designated to support the expanded activities related the AB1570 project. One position would support a proactive data quality program. Other projects expanding the Information Program would require additional staff.



- ⇒ **It is recommended that the Commission work with the Legislature and the Governor to provide the staffing required to attain and sustain a high quality and comprehensive Information Collection and Dissemination Program.**

### **Hardware and Software Replacement Cycle**

The Commission has been able to continue its Information Collection and Dissemination Program in spite of the budget cuts it has sustained only because it had a sound technological infrastructure in place. However, as the infrastructure has aged, it has required more maintenance. The time taken for this maintenance has come at the expense of the Information Program. Implementing a program to regularly update this infrastructure can alleviate this situation.

- ⇒ **It is recommended that the Commission establish a budget and schedule to update/replace its technology infrastructure.**

### **Authority to Collect Data**

While the Commission's enabling legislation specifies that it can request whatever data it needs to accomplish its mission, the reality has been that CPEC has no authority to mandate the collection of any specific data. The practical effect of this situation is that the Commission must either expend resources to negotiate the submission of data or find substitute data elements or sources. This situation impacts the ability of the Commission to rapidly gear up to deal with some emerging issues. The obvious solution is to provide the Commission with the recognized and accepted authority to ask for the data elements it requires. In this situation, the Commission has the obligation to request only those data that will be used for its work.

- ⇒ **It is recommended that the Commission request the segments to work with CPEC to develop interagency data-sharing mechanisms in order to secure the data required to fulfill its mission, and if that is unsuccessful within the next 30 days, the Commission work with the Legislature and the Governor to augment its authority to request such data.**

### **Enhancements to Data Collection System**

In order to create a more robust Data Collection System, the Commission needs to supplement its current set of data. The following recommendations highlight additions that should be made.

- ⇒ **Fully implement AB 1570 by adding a student identifier to all enrollment and completions records submitted by UC, CSU, and CCC.**

Adding a student identifier to all enrollment and degree records allows for longitudinal studies about student behavior. Currently, UC and CSU only provide data on students they identify as community college transfers for the year the student transfers. Typical studies could focus on time-to-degree, enrollment in multiple institutions during the same term and stop-out patterns.

- ⇒ **Collect enrollment data from UC, CSU, and CCC for the winter/spring terms.**

Currently CPEC only collects enrollment data for the fall term. While this provides a useful snapshot about student enrollment for a period of time, it only captures a part of the school year. Collecting winter and spring data will allow for a more complete picture of student behavior. In addition, our understanding of Community College Transfers will be enhanced since a number of community college students transfer to four-year institutions in the winter/spring terms. Better understanding about student behavior throughout the school year could improve planning efforts and enrollment management.

⇒ **Collect data on student applications, students admitted, and students who enroll at UC and CSU campuses.**

Currently CPEC only collects enrollment data on students who enroll at UC and CSU. Adding data on students who apply and students who are admitted but who don't enroll could help our understanding about student choices. This could allow for a better understanding of opportunities available for the college choices of students.

⇒ **Add Socio-Economic Status (SES) characteristics to the enrollment and completion data record already collected by CPEC.**

The Commission does not collect any SES data on students enrolled or students who complete their degrees. Collecting some SES characteristics would enhance our understanding about student make-up and behavior. These could include family income and parental educational attainment. Collection of SES data would provide a much better understanding about student options and choices (about college selection, ability to pay, the need to work, etc).

⇒ **Identify and collect data about workforce needs and workforce preparation.**

Collection of workforce data would allow the Commission to better understand the needs of California employers and be able to better evaluate the need for new programs, the continued existence of some programs, and provide information to high school students as they prepare to enter college and then the workforce. Data on workforce needs allows the Commission to make better policy recommendations in program review, the location of new college campuses, and provide better (targeted) information to students as they enter college and try to determine a career path.

⇒ **Request that the Independent Colleges and Universities (or at least the AICCU subset) submit individual enrollment and degrees data as do the public colleges and universities.**

Currently, the Commission collects aggregated data on students attending independent college and universities through the NCES data collection process. Collecting an individual record on every student enrolled and every degree awarded by an independent institution would allow the Commission to do comparable analysis with the public sector institutions and conduct longitudinal analyses of student cohorts.

## 8. Bibliography

*History and Scope of the California Postsecondary Education Commission's Information System Program*  
– Commission Agenda Item, April 27, 1987

*Issues of Data Collection About California's Independent Colleges and Universities – A Staff Report to the California Postsecondary Education Commission -- Commission Agenda Item – June, 1986*

*California's Experience with HEGIS/IPEDS Processing: A Candid Assessment* –Presentation at NCES IPEDS Conference – June, 1988

*Memo from Commissioner Stephen P. Teale forming a Committee to Evaluate the Effectiveness of Postsecondary Education and Final Report of Committee* – Commission staff report 1974-1975

*Feasibility Plan for a Comprehensive Student Information Study: A Report to the Legislature and Governor in Response to Assembly Bill 880 (1984)* – Commission Report 86-8, March 1986

John Harrison, The Wyndgate Group, Ltd. – *A Comprehensive Student Information System: A Report for the California Postsecondary Education Commission* – Report commissioned by CPEC, 88-32, September 1988

*Implementing a Comprehensive Student Information System in California: A Recommended Plan of Action* – Commission Report 88-36

### Internal Staff Papers

*Examples of Data Needed to Better Inform California Higher Education Policy and Funding Decisions* – Internal Draft

*Proposed Upgrades and/or Enhancements to CPEC Databases* – Internal Draft

*Summary Report of Current Commission Databases* – Internal Draft

*Scope of the Commission's Data Collection Enterprise* – Internal Staff Paper