

EDUCATION DATA IN CALIFORNIA: AVAILABILITY AND TRANSPARENCY

Janet S. Hansen

Senior Policy Researcher, The RAND Corporation¹

A paper prepared for the “Getting Down to Facts” Project

November 2006

Californians have a lofty ambition—educational excellence for all students—for their public schools and annually invest over \$45 billion² to achieve this worthy objective. How can policymakers and the public know how their schools are doing and whether their investments are accomplishing their goals? Do teachers have the information that would enable them to tailor their instruction to the specific needs of their students? Is information available to indicate which policies and practices show evidence of improving student achievement? Such questions are increasingly central to debates over education policy and the most effective way to allocate resources. The ability to answer these questions depends on strong data systems that collect the relevant information and make it available to various stakeholders in the education enterprise in accessible and understandable ways.

This paper examines the availability and transparency of K-12 education data in California and their capacity to support school improvement efforts and effective and efficient

¹ Affiliation included for identification purposes only. The opinions expressed in this paper are solely those of the author and do not represent those of RAND or any of its sponsors.

² In 2004-05 General Fund revenues for K-12 education from federal, state, and local sources totaled \$45.4 billion, as reported by Ed-Data (<http://www.ed-data.k12.ca.us/Navigation/fsTwoPanel.asp?bottom=%2Ffinance%2FFinanceReportForState%5FIntro%2Easp%3Flevel%3D04%26reportnumber%3D4%26fyr%3Dcurrent>).

use of education resources. Based on reviews of public documents and websites and interviews with twelve individuals knowledgeable about California and other state data systems and data system building, the paper concludes that California is lagging most other states in developing education data systems capable of helping policymakers and others understand how schools are doing and which programs are effective for improving student attainment and achievement. The state has only in the last several years begun to move beyond the traditional approach to data collection: emphasizing discrete, disconnected data “silos” that address reporting and monitoring requirements but do not lend themselves to analyses that can guide policy and program improvement. Recent steps to begin implementing longitudinal student and teacher data systems and to improve education management are promising, but the success of these undertakings is not yet assured. While efforts to build strong data systems undoubtedly encounter technological challenges, the primary barrier in California currently seems to be a lack of strong state advocates for the development of a first-class education data program.

The paper begins by describing how state and federal school reform efforts, federal data initiatives, and foundation-sponsored data improvement activities provide a context for considering data system development in California. It then provides an overview of the state of education data in California today and a discussion of the long way California still has to go to be capable of providing the information policy makers and citizens need to address key policy questions. Next, the paper examines the factors that appear to be holding California back in developing statewide K-12 data systems and illustrates what a “cutting-edge” education data system looks like by describing Florida’s data infrastructure and offerings. Finally, the paper poses challenges that California policy makers will need to address if the state is to successfully implement the new data initiatives it currently has underway.

THE CONTEXT FOR CONSIDERING DATA SYSTEM DEVELOPMENT IN CALIFORNIA

The roots of today's interest in data-driven decisionmaking can be traced to the standards and accountability movement of the 1980s and 1990s, which helped shift the focus of school reform from processes to educational outcomes. As states implemented new accountability programs, many also decided that they had to develop new or strengthened information systems capable of tracking student and school performance.

The federal No Child Left Behind Act (NCLB) of 2001 upped the ante by specifying numerous kinds of information that would have to be reported about students, teachers, schools, districts, and states, often in more detail (for example, subgroup test scores; "highly-qualified" teachers) than existing state and local information systems supported. Thus states like California which had not previously created integrated or longitudinal statewide information systems faced new pressures to do so in order to meet new requirements efficiently; states with existing statewide systems often found that their systems needed upgrading. In 2005 the federal government launched a new program, the Statewide Longitudinal Data Systems Grant Program, to support the development of state data systems. In November 2005 the Education Department's Institute for Education Sciences (IES) awarded three-year grants to 14 states (out of 45 that applied) to help them design, implement, and/or upgrade their systems. The California Department of Education (CDE) received a grant for \$3.255 million for a project commencing on May 1, 2006 to develop a longitudinal student information system. (Grants to additional states are expected to be made in subsequent years.)

Meanwhile, the federal government has also been moving to streamline its own data collections and move them to web-based reporting formats, a development with consequences for state data systems. In 2002 the U.S. Department of Education (ED) began the Performance-Based Data Management Initiative (PBDMI), an effort to consolidate 16 separate data collections (administered by seven program offices) into a single data collection system and to eliminate duplications, conflicting definitions, and information not needed to evaluate departmental programs. State education agencies are collaborating in this effort by helping to develop common data definitions and by piloting the new collection system while continuing to meet their ongoing reporting requirements.³ In addition to modifying data collection, the initiative involves the development of a web-based Education Data Exchange Network (EDEN) to provide state education agencies and the federal government the capacity to transfer and analyze information about education programs.⁴ Because of differences that remain to be resolved among participating ED offices and also because many states have not yet been able to provide the data through EDEN that would permit older collection mechanisms to be eliminated, PBDMI and EDEN are running behind their original implementation target dates.

The federal government is not the only outside influence on the development of state data systems. The National Education Data Partnership is a joint effort of several groups including the Council of Chief State School Officers, Standard & Poor's School Evaluation Service, the National Center on Educational Accountability, and the CELT Corporation. It is funded by the Bill & Melinda Gates Foundation and the Broad Foundation. Key activities include development

³ U.S. General Accounting Office, *Education's Data Management Initiative: Significant Progress Made But Better Planning Needed to Accomplish Project Goals*, GAO 06-6, Washington, DC: author, October 2005.

⁴ For further information about PBDMI and EDEN, see <http://www.ed.gov/about/inits/ed/pbdmi/index.html>.

of a robust web-based analytical service for education data (SchoolMatters⁵) and the provision of consulting and technical services to assist states in building their data infrastructures.

Another foundation-sponsored data-oriented activity is the Data Quality Campaign (DQC), created in 2005 with support from the Bill & Melinda Gates Foundation with the goal of seeing longitudinal student data systems established in all 50 states by 2009. Operated by the National Center for Educational Accountability, key DQC activities include promoting data systems through research studies and policy briefs, surveying the status of data system development in each of the 50 states, and providing snapshots of every state's progress in developing longitudinal student information systems. As will be indicated below, DQC surveys are an important source of information on how California's progress on developing longitudinal student information stacks up against the rest of the nation.

THE STATE OF EDUCATION DATA IN CALIFORNIA TODAY

If California has a data "problem," it is not a scarcity of data being collected and reported on students, teachers, and schools.

CDE currently⁶ has 125 active data collections. Key ones include:

- California Basic Education Data System (CBEDS): "The California Basic Educational Data System (CBEDS) is an annual data collection administered in October. The purpose of CBEDS is to collect information on student and staff demographics. There are three separate forms used to collect this data. The *County/District Information Form* is used to collect data specific to district and county offices on the number of classified staff, gifted and talented education, estimated number of teacher hires, high school graduation requirements, and student interdistrict transfers. The *School Information Form* is used to collect data specific to schools on the number of classified staff, school enrollment, high school graduates, enrollment in selected high school courses, career-technical education enrollment, dropouts, alternative education, technology, education calendars,

⁵ <http://www.SchoolMatters.com>.

⁶ As of July 1, 2006, according to CDE's online Data Resource Guide at <http://inet2.cde.ca.gov/dataresourceguide/>.

and No Child Left Behind (NCLB) reporting requirements. The *Professional Assignment Information Form* is used to collect data on certificated staff specific to district and county offices on their highest education level, racial ethnic designation, gender, birth year, educational service, course assignment, position, teaching credentials, and NCLB reporting requirements.”⁷ At one time the PAIF included individual staff salaries, but this is no longer the case.

- Fiscal data: local education agencies (school districts, county education offices, and “joint power agencies”) have since 2003-04 been required to report their unaudited actual end-of-year financial data to CDE using the Standardized Account Code Structure (SACS). SACS provides the framework for common reporting by defining codes for classifying every general ledger accounting transaction by the fund, resource, project year, goal, function, and object. SACS includes subcodes permitting school-level and not just district-level reporting, but subcode use is voluntary and data definitions that would be necessary to insure comparability of data at the school level (e.g., consistent allocation of centrally-provided services) have not been developed.
- Student achievement data: test vendors provide CDE with student scores on state tests, including the four components of the Standardized Testing and Reporting (STAR) program⁸ and the California High School Exit Exam (CAHSEE).

Additional data on teachers is gathered by the California Commission on Teacher Credentialing (CCTC, which is independent and not part of CDE) and the California State Teachers Retirement System (CALSTRS). CCTC collects data on the credentials teachers hold and which university recommends their credentials. CALSTRS collects data on when individual begin contributing to or withdrawing from the retirement fund. Universities and the state’s Employment Development Department also have information on teachers, such as their preparation programs and the wages paid to individual employees.

California makes much of the information it collects on students, schools, and teachers publicly available in various ways:

⁷ CBEDS description from the CDE website: <http://www.cde.ca.gov/ds/sd/cb/materials.asp>.

⁸ The four components of STAR are the California Standards Tests (CST), the California Alternate Performance Assessment (CAPA), California Achievement Tests (CAT/6 Survey - grades 3 and 7 only), and the Spanish Assessment of Basic Education, Second Edition (SABE/2).

- CDE has developed the **Data Resource Guide**, an online catalog of the department’s data products. The guide allows a user to access information about data collections, databases, and mandated reports. It indicates what data CDE collects, manages, and stores; where the data are located; who is the point of contact for the data; and how the data can be accessed and shared. As the website indicates: “Like a catalog in a library, the Data Resource Guide describes the CDE's data. It does NOT contain the actual data itself. Although all CDE data products are listed in the Data Resource Guide, some of the data may not be available to the public due to federal or state confidentiality laws.”⁹
- CDE also maintains an online system called **DataQuest** that provides information on accountability scores (e.g. California’s Academic Performance Index (API) and the Adequate Yearly Progress scores required by the federal No Child Left Behind Act), test data, enrollment, graduates, dropouts, course enrollments, staffing, and English learners. Data are available for various years at (depending on the data element) the state, county, district, and school level. The DataQuest site¹⁰ includes QuickQuest (a faster way of accessing data), a parents’ page to facilitate obtaining information about a specific district or school, methods for creating individualized reports, and links to sources of directly-downloadable data from CBEDS.
- **Ed-Data** is a website¹¹ that provides fiscal, demographic, and performance data on California’s K-12 schools. Pull-down menus provide access to data at the state, county, district, and school level. The site provides a glossary of terms relevant to understanding California’s education data and articles analyzing California education issues. Ed-Data is the product of a decade-old partnership among CDE (which makes data available), the Alameda County Office of Education (which designs, operates, and maintains the Ed-Data website), EdSource (an independent, not-for-profit organization which prepares graphic displays for the Ed-Data website, sets the context for the various reports and profiles, and monitors information on users and suggestions for future improvements), and the Fiscal Crisis & Management Assistance Team (an independent, state-funded entity which administers the partnership). The Ed-Data website is constantly being upgraded; the most recent (spring 2006) addition is district and county office of education financial data (provided using the common SACS accounting codes) showing how funds are raised and spent.
- California also provides data for **SchoolMatters**¹², a product developed by Standard & Poor’s School Evaluation Service to provide education data for all 50 states. The SchoolMatters website aims to provide state overview information as well as district and school-level information on such things as student performance; college preparation; NCLB; spending, revenue, and taxes; school environment; and community

⁹ <http://inet2.cde.ca.gov/dataresourceguide/>.

¹⁰ <http://data1.cde.ca.gov/dataquest/>

¹¹ <http://www.ed-data.k12.ca.us/>.

¹² <http://www.schoolmatters.com>

demographics. In addition, S&P calculates various ratios and indicators, such as a “return on spending index.” SchoolMatters is funded by the Broad and the Bill & Melinda Gates Foundations and is an outgrowth of the School Information Partnership announced in 2003 by President George Bush, which produced an NCLB-branded website of data on the 50 states with funding from the Broad Foundation and the U.S. Department of Education. SchoolMatters depends on data made available by state agencies as well as that collected from other sources; at this time not all states are choosing to provide data to the site.

- California has a **School Accountability Report Card (SARC)** which was mandated by Proposition 98 (The Classroom Instructional Improvement and Accountability Act) in 1988. SARC provides parents and the community with information about the condition and performance of each public school in the state. Subsequent state and federal laws have added content requirements for the SARC, which now include the following:¹³
 - State law
 - Demographic data
 - School safety and climate for learning information
 - Academic data
 - School completion rates
 - Class sizes
 - Teacher and staff information
 - Curriculum and instruction information
 - Postsecondary preparation information
 - Fiscal and expenditure data
 - Federal law
 - Adequate Yearly Progress of students
 - Title 1 program improvement status
 - Graduation rates at secondary level
 - Information about the extent to which “highly qualified” teachers are teaching core academic subjects

Senate Bill 687 (2005) expanded the existing content requirements of SARC by calling for new data to be added to the 2006-07 report card on salaries and expenditures, career technical education, and textbook sufficiency. For salaries, it requires each school to report the average of the actual salaries paid to teachers at that school. Per pupil expenditures are to reflect the actual salaries of personnel assigned to the school site.

CDE, which is responsible for most of the K-12 education data collected in California, has been working to manage its data programs more efficiently and effectively in response to

¹³ CDE, SARC Overview, August 2005, <http://www.cde.ca.gov/ta/ac/sa/documents/overview0805.ppt>.

both internal evaluations and to a 2002 consultant's report commissioned by the California Department of Finance (DOF). That report, conducted by MGT of America, Inc.,¹⁴ criticized CDE data management on a number of dimensions:

- Data collection within CDE was highly decentralized and varied within and between offices.
- There had been minimal efforts to coordinate data collection among CDE program offices and obtain a holistic picture of the data collected by the Department.
- Program offices managed data in a decentralized fashion without rigorous and aligned data management standards or a common vision about data management.
- There was no system for naming and defining data consistently throughout the department.
- There was limited and inconsistent data validation.
- The department relied heavily on paper submission of data.
- Data collections involved inconsistent units of analysis or inconsistent time periods.
- There was limited documentation of existing data.
- Data were stored in many different ways and locations.

The lack of common data definitions, the limited and inconsistent data validation, and the reliance on paper submissions compromised the accuracy and reliability of CDE data and inhibited data sharing across the department. Data sharing was also constrained by inconsistencies in units of analysis and time period coverage, by limited documentation, and by multiple approaches to data storage. Limited data sharing at CDE increased the reporting burden on districts, which had to respond to multiple data requests from various CDE offices.

CDE has taken a number of steps to respond to the problems identified in the MGT report (and in its own earlier internal evaluation). It established a Data Management Division with the responsibility of providing leadership throughout the department on managing data. The CDE staff have created "preferred data definitions" for 48 data elements, with 150 more planned over the next year, thus beginning the important process of standardizing data definitions throughout

¹⁴ MGT of America, *Data Management Study of the California Department of Education*, Sacramento, CA: author, April 2002.

various departmental data collections. Staff also developed the Data Resource Guide, a major contribution to making California's data collection instruments, databases, and reports more transparent to users. They have reduced data collections from 158 to the current 125, with hopes of further reductions. They paid visits to Florida and Ohio, states with reputations for strong data systems and plans, and have begun working on longitudinal data systems for both student and teacher information (more on these below).

CDE has also supported the work of the California School Information Services (CSIS) Program, a legislative initiative inaugurated in 1997 to develop and implement an electronic statewide school information system to facilitate the exchange and reporting of student information by school districts to CDE. CSIS operates independently of CDE; it is administered by the Fiscal Crisis and Management Assistance Team under the authority of the Kern County Superintendent of Schools but receives its funding from the state and must have a plan approved annually by the State Board of Education. CSIS provides a vehicle for collecting data extracted from the student information systems in each participating district. Instead of making aggregate reports to the state, districts report student-level data to CSIS that are then aggregated into required state reports as well as enabling the electronic transmission of student records among districts and to postsecondary institutions. CSIS requires participating districts to use standardized data elements and to use student information systems that meet CSIS standards for interoperability. Thus CSIS was originally envisioned as the vehicle through which California would develop comparable information systems in all districts, but participation was made voluntary, not mandatory. As of June 30, 2006¹⁵ 263 districts used CSIS to electronically transmit data required for three state collections: CBEDS, Language Census (R-30), and Student

¹⁵ Data from CSIS website: <http://www.csis.k12.ca.us/project-management/local-projects/CBEDS-List-of-Current-Program-Participants-FY0405-20060621.xls>.

National Origin Report (SNOR) reports. CSIS participants represent about half of student enrollments and about a quarter of school districts in the state.

Despite all the efforts to make education data available and understandable, the multiplicity of data sources in California can be confusing. The same CDE-collected data feed the various sites described above, but it not clear to a casual user what the distinctions are between the sites; and no road map exists to guide the user in making the choice about which source to consult. Users who know the data elements they want, for example, are probably better off beginning with DataQuest, while users whose starting point is an interest in a particular school or district will find Ed-Data more attuned to their needs. Users who are interested in financial data need to consult Ed-Data, which has recently added these data to the site; no financial data is available through DataQuest. On the other hand, DataQuest reports test score data, while Ed-Data reports API and growth scores but not scores on individual tests like CAHSEE or the tests in the STAR program

The SARC has become unwieldy, a fact that CDE recognizes. Superintendent Jack O'Connell announced in his 2006 State of Education remarks his intention to simplify the SARC:

“In this day and age of technology and innovation, we also have an obligation to better use data to not only drive our decision-making, but to more clearly and transparently articulate the successes and struggles of all our schools. Too often we talk in what I call, "educationese" -- a shorthand dialect of acronyms and jargon that is simply unintelligible to the average person. My friends, in order to succeed, we in the education community must stop speaking only to ourselves. It's time to put the public back in public schools. Along these lines I will begin this year by working to redesign our school accountability report cards in order to make them much more user friendly. These reports cards, that were intended to let parents and communities know how individual schools were doing, have become so unreadable that a UCLA study found them harder to comprehend than

several IRS forms and Microsoft Windows XP Driver Installation Instructions! I know we can do better than that.”¹⁶

CDE has redesigned the SARC template for the 2006-07 school year. Among other changes, it now includes a two-page “executive summary” to provide users with a quick snapshot of school accountability; more detail is available in the complete SARC. In September 2006, however, Governor Schwarzenegger vetoed a bill (SB 1510) supported by Superintendent O’Connell that would have further clarified and simplified the SARC.¹⁷

STILL A LONG WAY TO GO

Despite the plethora of education data available in California and efforts to improve the management and usability of those data, the state still has a long way to go to provide the data its policymakers and citizens need to understand how students and educators are performing and how resources might most effectively be allocated. For example, California currently lacks the ability to:

- Track students individually and over time and link them to performance scores and outcomes such as graduation.
- Track teachers individually and link them to students taught, to their preparation programs, and to the professional development they receive.
- Provide teachers with student histories and performance indicators.
- Link resource use and performance.

Moreover, the state has only fragmentary efforts underway to link K-12, postsecondary, and post-school records, therefore severely hampering efforts to understand what happens to students and why as they move through the educational system and into the workforce.

¹⁶ “State of Education 2006,” prepared remarks by State Superintendent of Public Instruction Jack O’Connell, February 7, 2006, <http://www.cde.ca.gov/eo/in/se/yr06stateofed.asp>.

¹⁷ Governor Schwarzenegger’s veto message is at http://gov.ca.gov/pdf/press/sb_1510_veto.pdf. Superintendent O’Connell’s press release on the governor’s veto is at <http://www.cde.ca.gov/nr/ne/yr06/yr06rel121.asp>.

To illustrate how the absence of such capabilities inhibits decisionmaking, the Data Quality Campaign has identified six “priority questions” facing policymakers and educators and indicated how many states have data systems capable of answering them.¹⁸ The number of states capable of answering each question is shown below in parentheses; California is not capable of answering any of the questions with its currently-available data:

- Which schools produce the strongest academic growth for their students? (23 states)
- What achievement levels in middle school indicate that a student is on track to succeed in rigorous courses in high school? (5 states)
- What is each school's graduation rate, according to the 2005 National Governors Association graduation compact? (26 states)
- What high school performance indicators (e.g., enrollment in rigorous courses or performance on state tests) are the best predictors of students' success in college or the workplace? (4 states)
- What percentage of high school graduates who go on to college take remedial courses? (14 states)
- Which teacher preparation programs produce the graduates whose students have the strongest academic growth? (9 states)

While education data were historically collected on an annual basis for administrative purposes, it is increasingly recognized that program improvement requires the development of longitudinal databases on students so that individuals' progress can be tracked over time. The DQC, in support of its goal to foster the development of longitudinal data systems in every state, has identified ten essential elements of such systems:

1. A unique statewide student identifier.
2. Student-level enrollment, demographic, and program participation information.
3. The ability to match individual students' test records from year to year to measure academic growth.
4. Information on untested students.
5. A teacher identifier system with the ability to match teachers to students.
6. Student-level transcript information, including information on courses completed and grades earned.
7. Student-level college readiness test scores.
8. Student-level graduation and dropout data.

¹⁸ See http://www.dataqualitycampaign.org/survey_results/policy.cfm.

9. The ability to match student records between the PreK-12 and higher education systems.
10. A state data audit system assessing data quality, validity, and reliability.

DQC has surveyed the states annually for the past several years to monitor how states are doing in implementing these elements. In 2005,¹⁹ California ranked near the bottom, with only one of the elements (information on untested students; a second element, the assignment of a unique student identifier to every student in the state, was about to be completed). In the just-released 2006 survey,²⁰ California still ranks below average, although it now has four of the elements (information on untested students; a unique student identifier; student level enrollment, demographic, and program participation information; and a state data audit system).

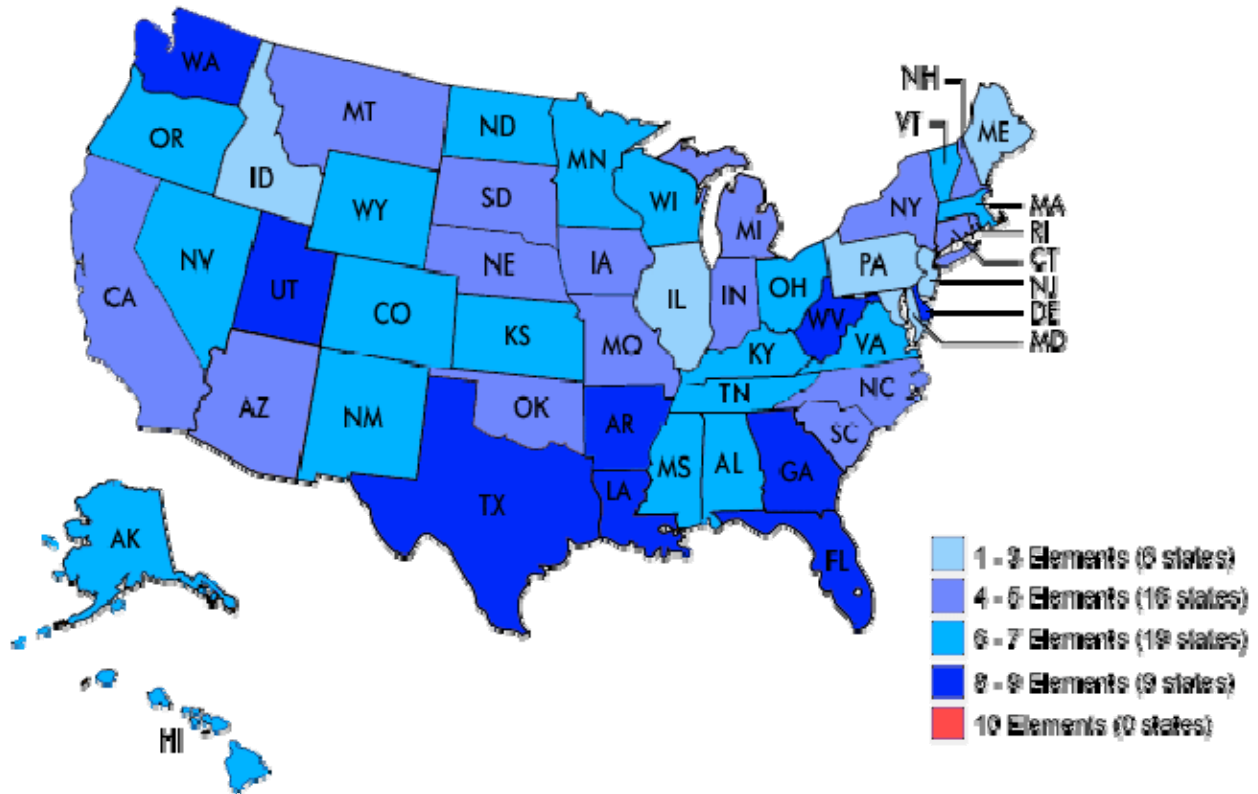
[See the figure on the following page]

¹⁹ http://www.dataqualitycampaign.org/activities/survey_result_2005.cfm.

²⁰ http://www.dataqualitycampaign.org/survey_results/index.cfm.

DATA QUALITY CAMPAIGN 2006 SURVEY

How Many of 10 Essential Elements of a Longitudinal Data System Does Each State Have?



For its 2006 Technology Counts report,²¹ *Education Week* also surveyed states, among other things inquiring whether states had trackable ID systems for teachers. Forty-three states

²¹ Technology Counts 2006, "The Information Edge: Using Data to Accelerate Achievement," *Education Week*, Vol. 25, Issue 35, May 4, 2006.

reported having such systems; California does not. Some states not only have such systems but have developed the capacity to link student and teacher data.²²

California has several data initiatives underway that promise to fill some of the gaps in its current education data capabilities, but as presently envisioned, even when (or if) completed, they will leave some significant holes.

One major initiative is CALPADS—the California Longitudinal Pupil Achievement Data System. Authorized by SB 1453 in 2002 and SB 257 in 2003, CALPADS will enable California to comply with the requirements of No Child Left Behind. The authorizing legislation calls for a data base “to assess the long-term value of [the state’s] educational investments and programs and provide a research base for improving pupil performance.” SB 1453 authorized the assignment of unique student identifiers, a process that is nearly complete (a few districts are holding out). A Feasibility Study Report for CALPADS was completed by CDE in August 2004 and approved by the Department of Finance in October 2005. The RFP is scheduled to be completed and approved in 2006, and a vendor is to be selected in July 2007. The CALPADS system is scheduled for completion in December 2008.

At the insistence of the California Department of Finance (perhaps among others, though not the authorizing legislation), CALPADS will include only those data elements required by NCLB. This will leave data of potential interest to education policymakers and researchers, including some connected to other federal programs such as vocation education and special education, out of CALPADS and will limit the ability of users to assess the value of educational

²² The 2006 Data Quality Campaign survey indicates that 16 states (Arkansas, Delaware, Florida, Georgia, Hawaii, Kentucky, Louisiana, Mississippi, New Mexico, Ohio, Rhode Island, South Carolina, Tennessee, Utah, West Virginia, and Wyoming) have this capacity. Asking the question somewhat differently and using more criteria for having both “basic” and “advanced” data, *Education Week* in 2006 determined that five states (Arkansas, Georgia, Louisiana, Ohio, and Tennessee) had “advanced data systems” for both students and teachers and the ability to link data for the two groups.

investments and programs. Current plans for the data system restrict its elements to those required by NCLB Title I Part A (Education for the Disadvantaged), Title II Part A (Preparing, Training, and Recruiting Highly-Qualified Teachers), Title III Part A (English language acquisition), Title IV Part A (Safe and Drug-Free Schools and Communities), and Title IX (General Provisions, relating to expulsions).²³ As currently envisioned in the feasibility study, “CALPADS will not be a ‘data warehouse.’ That is, the system will not maintain the detail and summary aggregations of student-level transactions in a manner specifically designed to support elaborate and complex data selection queries for decision support analysis and research.”²⁴ The study does note that “CALPADS data may serve as source data for a data warehouse or decision support system, should one be developed.”²⁵

First steps are also being taken on a longitudinal teacher data system to bring together data currently housed at CDE, CCTC, and EDD. Legislation in 2005 (SB 77, the 2005-06 budget act) authorized a feasibility study on a longitudinally-linked teacher data system with the single stated intention of supporting high-quality program evaluations, especially for state investments in teacher professional development and retention programs aimed at improving teacher quality. The feasibility study envisions a Teacher Data System (TDS) including all certificated staff (not just teachers) that will:

- Provide the data necessary to assess the effectiveness of policies, programs, and expenditures.
- Provide the data necessary to evaluate teacher mobility, retention, and attrition.
- Provide the data necessary to evaluate the pathways to teaching.

²³ California Department of Education, Technology Services Division, *California Longitudinal Pupil Achievement Data System (CALPADS) Feasibility Study Report*, August 20, 2004, available at http://www.educationroundtable.org/PowerPoints_Handouts/Sept9MeetingHandouts/Tyler/040820_FSR_Final.pdf.

²⁴ *Ibid.*, p. 6.5.

²⁵ *Ibid.*, p. 5.7

- Provide the data necessary to monitor teacher assignments.
- Provide an efficient, flexible, and secure means of maintaining longitudinally-linked statewide teacher level data.
- Promote good data management practices with respect to teacher data systems and issues.²⁶

The feasibility study for the TDS was completed in March 2006 and approved by the Department of Finance in May 2006, with the proviso that, in terms of integrating CDE teacher data into the TDS, only the teacher data in CALPADS is covered by the DOF approval. Therefore, “any activity linking *teacher* data with *student* data must be addressed through a separate proposal” (emphasis added).²⁷ The 2006-07 Budget Act (AB 1801, approved in June 2006) appropriates funds for “the Department of Education and the Commission on Teacher Credentialing to begin development of the system by issuing a request for proposals and by beginning initial data management activities necessary for successful implementation of the database.”²⁸ CDE and CCTC are expected to develop a system of unique teacher identifiers during 2006-07.²⁹

In August 2006 the Legislature passed SB 1614, a bill calling for a “comprehensive state education data information system in the department [CDE] that includes information regarding the teacher workforce.” The bill includes within the California Education Information System both CALPADS and a California Longitudinal Teacher Integrated Data Education System (CALTIDES) based on the feasibility study for the TDS. The legislation appears to open the door to the possibility at some future point of linking student and teacher data, a linkage opposed

²⁶ California Department of Education, Data Management Division, *Teacher Data System (TDS) Feasibility Study Report*, March 31, 2006.

²⁷ Letter from the Department of Finance to the Superintendent of Public Instruction, May 12, 2006, Log Number 2006-380.

²⁸ CA state budget 2006-07, <http://www.ebudget.ca.gov/pdf/Enacted/BudgetSummary/FullBudgetSummary.pdf>.

²⁹ Memo from Sue Stickel, Deputy Superintendent, Curriculum and Instruction Branch, CDE to State Board of Education, July 7, 2006, <http://www.cde.ca.gov/be/ag/ag/yr06/documents/blue-jul06item08.doc>.

(according to several interviewees) by the Department of Finance and by the powerful California Teachers Association (CTA). Despite its historical opposition to a teacher data base and to linking student and teacher data, SB 1614 garnered CTA support, according to interviewees, because it contains what the union apparently views as a “firewall” against using education data for teacher evaluation or pay: “data in the system may not be used, either solely or in conjunction with data from the California Longitudinal Pupil Achievement Data System, for purposes of pay, promotion, sanction, or personnel evaluation of an individual teacher or groups of teachers, or of any other employment decisions related to individual teachers. The system may not include the names, social security numbers, home addresses, telephone numbers, or e-mail addresses of individual teachers.”

California has two initiatives underway that begin to bridge the gap between its K-12 and postsecondary data systems, but they are both limited and largely involve (at the postsecondary level) the community college system rather than the state colleges or universities. Linkages are hampered by the fact that there is no unitary data system for California’s public higher education institutions, with each of the three sectors (California Community Colleges, California State University, University of California) having its own policies and practices for information collection and dissemination.

Cal-PASS (California Partnership for Achieving Student Success) is a voluntary data-sharing arrangement among consortia of California elementary schools, secondary schools, and colleges and universities. Currently there are ten of these consortia acting as regional student tracking partnerships. Cal-PASS started in 1998 as a local project in San Diego and Imperial counties to foster sharing of student data between K-12 schools, community colleges, and universities. It has received financial support from the California Community Colleges system

and now receives state funding. Would-be participants sign memoranda of understanding agreeing to share data on student demographics; attendance; enrollment; course performance; test scores (STAR and high school exit exam); diplomas, degrees, and certificates; and (optionally) program interventions.

By joining a regional consortium, Cal-PASS participants get access to:

- A web-accessible, password-protected site which generates reports on the performance of students attending schools within the consortium and allows authorized users to explore aggregated data. These reports may be generated by institution, by year, or and/or by subject area.
- A web-accessible, password-protected site where unitary record level data may be downloaded by authorized users in a standard format that can import to a variety of analysis tools. This allows users to perform analysis at a deeper level than the standard reports.
- Access to Cal-PASS user group meetings and user web portal to share ideas about uses of Cal-PASS data.

Cal-PASS contains important elements of a higher education student record system that links to students' K-12 experiences, but its fragmented regional structure and incomplete statewide coverage makes it limited as a tool for state-wide policymaking.

The California Community Colleges system is developing a statewide internet-based system (CCCTran) for requesting, viewing and transmitting academic transcripts among authorized educational institutions and their trading partners. "Beta testing" of an electronic transcript system involving live exchanges of data among 8 community colleges and 6 state universities is currently underway. Eventually links between the K-12 and community college system are planned; but again, these appear to be likely to develop on a fragmentary and voluntary basis.

WHAT SEEMS TO BE HOLDING CALIFORNIA BACK IN DEVELOPING STATEWIDE K-12 DATA SYSTEMS?

As the Data Quality Campaign and *Education Week* survey results demonstrate, California is lagging many other states in developing statewide data systems, especially longitudinal ones, that can inform school improvement efforts. Despite recent progress, individuals interviewed for this paper expressed skepticism about whether the current initiatives would be carried out as envisioned or on time. What seems to be holding the state back?

One factor appears to be the absence of a “culture of data” in California that emphasizes the necessary connection between good data and school improvement efforts. Policymakers in some other states recognized and acted on this link many years ago. Thus, Florida, for example, began developing common data definitions for education data as early as 1976 and was a pioneer in developing statewide databases when it began planning for its current information system in 1984. Today, its education data system (now enhanced by a data warehouse that including so-called P-20 longitudinal coverage of students and teachers), is widely recognized as the most extensive in the nation. (Florida’s system will be described further below.) The development of high-quality statewide education data in Florida was fostered by a long-time legislative emphasis on making data-driven decisions and by legislative mandates requiring (and funding) the creation of robust data systems. Similarly, lawmakers in Texas included information systems in the major 1984 school reform legislation that resulted from the recommendations of the “Perot Committee.” They mandated creation of a Public Education Information Management System (PEIMS) to collect and make available information on schools and students; PEIMS (which collected data for the first time in school year 1987-88) promoted data quality by defining what data were to be reported by school districts and in what formats. In 2001 the Legislature began

funding the Texas Prekindergarten-16 Public Education Information Resource (TPEIR), a collaboration among the state's education agency, higher education coordinating board, and teacher certification agency to build an integrated data repository that would provide a comprehensive view of public education in Texas.

A "culture of data," was probably easier to foster in states such as Florida and Texas with comparatively strong traditions of state direction of education, in contrast with California's greater emphasis on local control. Florida also has the advantage of having only 67 school districts (though Texas has over 1,000). It is worth noting, however, that Iowa, a state with local control traditions as strong or stronger than California's (and with a large number—371—of school districts to accommodate), has in recent years implemented Project EASIER (Electronic Access System for Iowa Education Records) and created a statewide student identifier system and a teacher identifier system. State and area education agency personnel are now discussing the desirability of augmenting the state's "live" information systems with a data warehouse of cleaned, stable data to enhance reporting and enable analyses of education issues such as the impact of mobility on student learning and the impact of early intervention services.

A corollary to the absence in California of a state-level emphasis on the importance of education data has been the absence of incentives for school districts to care about the substance and quality of the data they provide to the state. In California, data flows up from districts through all the data collections that were mentioned earlier, but very little (with the notable exception of student test score data) flows back down to them except via the School Accountability Report Card (and its shortcomings have already been noted). In Florida, by contrast, there has been a strong emphasis on building local support for statewide data systems by taking seriously a *quid pro quo*: those who supply data need to get something back.

Therefore, Florida officials have used statewide data collections to reduce the burden on local officials, to foster the exchange of data among districts and between districts and higher education institutions, to give teachers data they can use to improve classroom instruction, and to provide feedback to high schools and community colleges on how their students have fared as they moved to the next educational level. These services will be further described below.

A number of individuals interviewed for this paper also indicated that the development of statewide data systems in California was being held back by reservations at the Department of Finance (DOF), which is responsible for the budgeting and control of information technology expenditures throughout state government.

Some of this hesitancy is understandable, given plenty of evidence from the business sector as well as from public agencies at the state and local level that big data projects not infrequently go seriously and expensively awry.³⁰ DOF has long had concerns about CDE's data management capabilities, as illustrated by its commissioning of the 2002 outside review by MGT of America.

Moreover, provisions of the California state constitution barring unfunded mandates have made DOF and other state officials wary of imposing data requirements that could cause school districts to insist on the need for state funds to modify their local information systems.

According to the constitution:

³⁰ An Organisation for Economic Cooperation and Development (OECD) policy brief cites estimates that fewer than a third of government and industry IT projects in 2000 were successful with regard to budget, functionality, and timeliness; almost a quarter were cancelled (OECD, "The Hidden Threat to E-Government: Avoiding Large Government IT Failures," PUMA Policy Brief No. 8, Paris, 2001). The Washington Post (August 18, 2006) recently reported on "The FBI's Upgrade That Wasn't: \$170 Million Bought an Unusable Computer System." Closer to home, North Carolina has recently encountered a number of problems in replacing a 20-year old student information system with a new data system called NCWise, and Idaho shut down its effort to build the Idaho Student Information System after encountering large unexpected cost projections. Idaho's experience has been analyzed in a new report from the Idaho Legislature's Office of Performance Evaluations, "Idaho Student Information Management Systems (ISIMS)—Lessons for Future Technology Projects," Report 06-02, August 2006.

Whenever the Legislature or any state agency mandates a new program or higher level of service on any local government, the State shall provide a subvention of funds to reimburse that local government for the costs of the program or increased level of service...³¹

Exceptions to this unfunded mandate provision are provided for

appropriations required to comply with mandates of the courts of the federal government which, without discretion, require an expenditure for additional services or which unavoidably make the provision of existing services more costly.³²

In 2001, Governor Gray Davis vetoed a bill that would have required CSIS to be fully operational in 90 percent of school districts and county offices of education by 2004-05, in part because of concerns about the potential state mandate costs that might result.³³ These constitutional provisions explain at least in part why DOF has insisted that CALPADS be limited to data elements required by the federal NCLB. (They also apparently help explain why the SARC became so unwieldy and unusable, since it was authorized by a ballot proposition and not by the legislature and thus became a convenient vehicle for imposing data requirements while skirting the constitutional unfunded cost requirement.)

Despite these legitimate concerns, however, some supporters of an improved education data structure for California view the DOF as excessively cautious and a stumbling block to improving the system. The Legislative Analyst's Office (LAO) has publicly commented in the past on the delays at DOF (and at the Department of Information Technology) in approving the scope of work for a project related to CSIS.³⁴ It took DOF 14 months (from August 2004 to

³¹ California Constitution Article 13B Government Spending Limitation Sec. 6(a).

³² California Constitution Article 13B Government Spending Limitation Sec. 9.

³³ California Legislative Analyst's Office, Analysis of the 2002-03 Budget Bill, http://lao.ca.gov/analysis_2002/education/ed_13_CSIS_anl-02.htm.

³⁴ California Legislative Analyst's Office, Analysis of the 2001-02 Budget Bill, http://www.lao.ca.gov/analysis_2001/education/ed_12_Databases-anl01.htm.

October 2005) to approve the feasibility study report for CALPADS, without which CDE could not move ahead to preparing an RFP for the system. (In a perhaps hopeful sign, DOF took only 2 months to approve the feasibility study report for the TDS in 2006.) Individuals interviewed for this paper (not CDE employees, it should be noted), commented on the long-standing “bad blood” between DOF and CDE and on the “hundred ways of saying no” that DOF has used to slow down CALPADS development.

Most recently, DOF and CDE have been at odds on the most effective way to fund local districts to upgrade and maintain their capacity to provide reliable student-level data (including student identifiers). CDE and LAO wanted funds included in the 2006-07 budget to provide “CALPADS incentive grants” to districts to support local data activities and compensate for the work involved in maintaining the new student identifier system, the quality of which will be essential to CALPADS’ successful implementation. In exchange for funding, CDE would have required districts to meet several quality measures related to the identifiers. The department wanted incentives of \$5 per pupil; it and the LAO supported provisions in the Senate and Assembly budget bills that would have provided \$2.50 per pupil, costing \$15 million. These funds were removed from the budget act right before it was finally approved; apparently DOF believes that funding for maintaining student identifiers and for other data quality enhancements related to CALPADS need not be provided to districts until CALPADS is operational.

This latter disagreement, finally, seems reflective of a more general issue: as one interviewee described California’s efforts to improve its education data systems, “the stakeholders assert they are supportive, but they’re really not invested in making it happen.” Original projections for CSIS envisioned a 5-6 year development period for full deployment to all districts, but the anticipated funding to reach that objective was not provided. A legislative

analyst described the longitudinal student data system now represented by CALPADS as a system “that’s been ready to finish in two years” for the last eight years. Despite recent progress on the student and teacher longitudinal data systems, if California continues to approach the development of its education data infrastructure in such a half-hearted fashion, it will miss out on the enormous potential that modern technology offers for using data to enhance teaching and learning. What that potential is can be seen by looking at how one “cutting edge” state has been able to capitalize on a belief in and commitment to good education data.

FLORIDA: WHAT A “CUTTING-EDGE” EDUCATION DATA SYSTEM LOOKS LIKE

Thanks to thirty years of attention to improving its education data infrastructure, Florida currently has a comprehensive K-20 data system covering public pre-Kindergarten through graduate school and containing student-level data for public schools, community colleges, career and technical education, adult education, and the state university system. The state can link student information to post-school employment and non-education program data. Longitudinal information from separate data systems are increasingly linked through an education data warehouse, and data analyses likely to be of wide interest are made available through data marts. State officials say that the development of this system has depended in particular on common course numberings and catalogs throughout the state, state academic standards for students, a common K-20 student identifier, and a state culture of data sharing.³⁵

³⁵ See, for example, the presentation by Hanna Skandera, Deputy Commissioner of the Florida Department of Education, at the National Governors Association February 2, 2006 data summit, <http://www.nga.org/Files/pdf/0602DATASUMMITSKANDERA.PDF>.

Florida continues to expand its data offerings through new initiatives (such as Sunshine Connections for teachers, performance profiles for legislators, high school and community college feedback reports, and a counseling tool for students) that bring the benefits of high-quality information to virtually all the stakeholders in the education system.

Origins: FIRN and the Development of K-12 Data Bases

Florida's education data journey³⁶ began in the 1970s with legislative initiatives aimed at improving program accountability and the speed and efficiency of data exchange between the state Department of Education (FLDoE) and school districts. The result was the creation of the Florida Information Resource Network (FIRN) which provided the "backbone" connecting FLDoE and districts electronically. All districts had access to FIRN by the end of 1985. Over the next few years state administrators gradually phased in the current K-12 data collection systems. As pioneers, they had no models to follow. In 1986, FLDoE established state-level student, staff, and finance data bases. Those districts with the technological capability reported finance data electronically in 1986; automated student- and staff-level data reporting began in 1987. In 1988 FIRN added the capability of transferring permanent student records and transcripts between schools and postsecondary universities. By 1991 all districts were automated, and state funding was awarded solely on the basis of information in the automated

³⁶ This portrait of the development and current status of Florida's education data system is taken from the following sources: Florida's application for a longitudinal data system grant from the U.S. Department of Education (<http://nces.ed.gov/Programs/SLDS/PDF/Florida.pdf>); presentations by Florida representatives at the National Governors Association February 2, 2006 data summit (<http://www.nga.org/portal/site/nga/menuitem.9123e83a1f6786440ddcbeeb501010a0/?vgnextoid=8a9fd3c900c09010VgnVCM1000001a01010aRCRD>); Hanna Skandera's presentation from the summit (see the previous footnote); the Data Quality Campaign's case study on building Florida's data system (http://www.dataqualitycampaign.org/files/State_Specific-Florida_2006_Site_Visit.pdf); and a 2004 presentation on the Education Data Warehouse prepared by the Florida Department of Education (<http://www.wiche.edu/Policy/WCALO/ppt/Sellers.ppt>).

student database (providing “high stakes” that provided a strong incentive for districts to care about data quality). Summary data collections were discontinued. State budgets currently require that a portion of the funds allocated to school districts must be used for data and information systems.

Florida’s student database links to its staff and finance databases. The staff database (which includes all personnel including school board members and janitors) can in turn be linked to certification files and contains information on demographics, compensation, experience, degree level, and other types of data.

FETPIP

In 1988, Florida began development of the Florida Education and Training Placement Information Program (FETPIP). This cooperative effort across state agencies links schooling and post-school data. It integrates many administrative data sources of state and federal agencies and is designed to follow up on all participants in Florida education and training programs at all levels to find out what happens to them after they leave these programs. Examples of student outcome data collected in FETPIP include in-state postsecondary enrollment, out-of-state postsecondary enrollment, occupational licensure, federal employment, public assistance, incarceration, military enlistment, and Florida employment and unemployment. Cooperating agencies include a host of state agencies as well as the U.S. Postal Service, the Department of Defense, the Office of Personnel Management, the U.S. Census Bureau, and the National Student Clearinghouse. FETPIP operates on the principal of *quid pro quo*: those who supply the data need to get something back. So, for example, FETPIP works with Florida’s welfare agencies to conduct staged follow-up on participants during and after their involvement in welfare programs

to monitor the reduction in welfare that occurs as participants become increasingly employable.

FETPIP data are used by public schools, vocational institutions, community colleges, universities and the workforce developmental system in their performance measurement programs.

Florida Education Data Warehouse

Despite the strengths of these various databases, they reflect their origins as administrative rather than analytical tools. As such, they had many limitations. Operational systems were isolated by business functions. K-12 and higher education data were not integrated. Multiple data and technical architectures made data linkages difficult. Readily-available data were limited.

In the late 1990s Florida began planning for an Education Data Warehouse with the following objectives:

- Complete, timely, accurate data
- Statewide view
- Data from multiple sources and organizations
- Integrated technical environment
- Historical data merged with current in a structured repository
- Comprehensive data definitions
- Cross-functional data management and support group
- Easy access and manipulation

The warehouse was developed over three to four years at a cost of about \$7 million (not including ongoing support and maintenance which requires 12 staff). The warehouse is a repository for cleaned, stable data going back to 1995 that links records in a student-centric, longitudinal, integrated fashion that makes it possible to follow Florida students from pre-Kindergarten through postsecondary education (if they attend in Florida) and into the workforce. The warehouse includes information on students, staff, educational institutions, curriculum, courses taken, and facilities. The warehouse produces pre-defined reports, data marts (subsets of

data tailored to answer a specific question or support a specific business need), and “metadata” (web-based, user-friendly access to data definitions and documentation). As of early 2006 the warehouse contained information on over 10 million students and nearly a million staff.

The Florida Education Data Warehouse is an evolving structure. In 2005 the state successfully applied for a \$1.6 million grant from the U.S. Department of Education’s Institute for Education Sciences to support further development of the warehouse. Specifically, between 2006 and 2009 Florida will incorporate school finance data and additional information on facilities into the warehouse. The state expects that this “will provide an unprecedented, new capacity to tie indicators of student progress and outcomes directly to detailed administrative, instructional, and program-level costs at the classroom level. Similarly, indicators of student success will be linked to details about the facilities in which they are educated.”³⁷ As part of the grant, the state promised to produce a guidebook that will enable other states to learn from Florida’s experience in establishing seamless integrated student, finance, and facilities information systems.

What does Florida do with all these data?

FLDoE cites many examples of how its data are used for accountability, evaluation, data-driven decisionmaking, and research:³⁸

Accountability

- School Grades
- Adequate Yearly Progress
- Postsecondary Performance Funding
- Postsecondary Accountability Reporting
- Common Core Measures for Workforce, Adult Education

³⁷ IES application abstract?

³⁸ Presentation to 2006 data summit

Agency “Performance Dashboards”

Evaluation

- Scholarship Programs
- Just Read, Florida! and Reading First
- Return on Investment
- Cost benefit analyses

Data-driven decisionmaking

- Third Grade Progression
- Pay for Performance
- Course taking patterns
- Strategic Planning

Research

- Teacher Effectiveness
- Charter Schools
- Achieving the Dream
- Articulation
- Labor market performance adjustments
- Longitudinal Employer Dynamics-U.S. Census
- Economic Impact Analyses

To flesh out just a couple of these examples, the availability of longitudinal achievement data for students enables the state to base its accountability system on student growth, not merely on attainment of specified levels of achievement. Student-level data held by the state facilitates rapid exchange within and between levels of education (as when students move across districts or apply to Florida colleges). The ability to link student, staff, and finance data bases has enabled FLDoE to undertake analyses of program effectiveness and to create a “Return on Investment” website³⁹ allowing users to evaluate measures of performance in light of the resources allocated to districts and schools.

³⁹ <http://roi.fldoe.org>.

FLDoE staff themselves use the data for research and analytical purposes. Other users include the state Office of Program Policy Analysis and Government Accountability and researchers from universities and other research organizations.

Several ongoing initiatives indicate how FLDoE works proactively to make its education data available and useful for a variety of stakeholders.

Legislators, for example, are provided with customized performance profiles on students and schools in their districts, with statewide data included for comparison. Data from the written reports are also accessible and searchable on the web.⁴⁰ Future reports will focus on postsecondary education and on other key indicators.

High schools and community colleges receive “feedback” reports⁴¹ with data on a variety of pre-graduation and post-graduation indicators, again with district and statewide information for comparison. These help schools and institutions learn about the progress of their students at the next educational level.

Students have since 2000 had access to facts.org,⁴² an online counseling tool that helps them monitor their progress towards educational goals. Via the web students can determine career objectives, evaluate high school progress, see high school course summary and grades, learn about higher education opportunities in Florida, apply to college online, get help choosing the right major, access college transcripts and grades, track progress towards college graduation, and learn about research indicating what matters in bachelor’s degree completion.

⁴⁰ <http://data.fldoe.org/performanceProfiles/K12/>.

⁴¹ <http://data.fldoe.org/readiness/default.cfm>.

⁴² <http://www.facts.org/cgi-bin/eaglec?MDASTRAN=SW-HMMNE00>.

Teachers are the targets of Florida’s latest major data initiative, Sunshine Connections.⁴³ Through a partnership with the Microsoft Corporation, Florida will over five years provide all its teachers with desktop, immediate access to classroom management tools, student performance data, instructional strategies, collaboration and communication abilities with other teachers, curricular materials, and personalized professional development opportunities. A public area on the Sunshine Connections website provides all users with a series of free tools organized by area of interest; a restricted area (still under development) will provide teachers with confidential tools and information specific to their own students. In an especially creative use of its “data pioneer” status, Florida attracted 13 bids from companies willing to partner on Sunshine Connections and to pay to develop tools which they can then market to other states.⁴⁴ Microsoft has committed \$6 million to the effort over five years.⁴⁵ Florida negotiated perpetual free licenses for access to the tools that will be provided through Sunshine Connections.

Researchers have obtained microdata from FLDoE to conduct a number of studies of interest to the state, including analyses of third grade progression (Harvard University), school choice (Manhattan Institute), and teacher effectiveness (researchers from six universities around the country).

⁴³ http://www.sunshineconnections.org/home_files/home_status.htm.

⁴⁴ Commissioner John Wynn cited this number in his presentation to the National Governors Association February 2, 2006 data summit; see <http://www.nga.org/portal/site/nga/menuitem.9123e83a1f6786440ddcbeeb501010a0/?vgnnextoid=8a9fd3c900c09010VgnVCM1000001a01010aRCRD>.

⁴⁵ Microsoft press release, Aug. 19, 2005, <http://www.microsoft.com/presspass/press/2005/aug05/08-19SunshineConnectionsPR.msp>.

CHALLENGES FOR CALIFORNIA

California, as all states, will have to develop an education data infrastructure compatible with its unique circumstances and needs. While a visionary system like Florida's can provide inspiration for discussions about what the state's ultimate objectives might be, the immediate task ahead is to effectively pursue the constructive initiatives already underway: CALPADS, CALTIDES, better linkages between the K-12 and higher education sectors, and improved data management through continuing efforts at CDE to develop common data definitions, reduce data collections where possible, and make data easier for localities to provide and more accessible and transparent for users. Though California's late start on some of these efforts will keep it behind more pioneering states, one clear lesson from past failures of major IT projects is, as a veteran of North Carolina's difficulties implementing a new statewide information system put it, "don't bite off more than you can chew."⁴⁶ This is especially true given what still needs to be learned; one of IES's goals for its longitudinal data grants program is the identification of best practices in "designing, developing, implementing, and utilizing statewide longitudinal data systems."⁴⁷

California needs to address several interrelated challenges if its current data system "bites" are to be successfully swallowed and digested.

Leadership: California needs influential education data "champions." It seems unlikely that the state can build data systems capable of supporting data-driven policy and funding decisions without strong and long-term support from state leaders who can reduce inter- and intra-departmental and inter-sectoral rivalries and ensure that funding and other necessary support is available. Decisions such as how extensive CALPADS should be (i.e., limited to NCLB required elements or incorporating other data desired by California educators and policymakers) deserve wider discussion than they are now getting.

⁴⁶ Sam Dillon, "States Struggle in Efforts to Computerize School Records," *The New York Times*, May 15, 2006.

⁴⁷ Cited in the U.S. Department of Education RFP for statewide longitudinal data system grants; <http://www.ed.gov/programs/slds/2005-372.pdf>.

Funding. One way that lukewarm support for education data system development in California has manifested itself is in inadequate funding to carry out system implementation. CSIS was originally intended to be deployed to all school districts over a five-to-six year period, but inadequate funding prevented the full roll-out. In January 2006 CSIS staff estimated that \$42 million would be needed to fund all the remaining districts at the “legislated compensation rate” normally given to help them fully participate in the system.⁴⁸ Instead, \$31 million was budgeted for what an Assembly budget subcommittee dubbed “CSIS light”: a less intensive version to help nonparticipating districts with hardware and software improvements necessary to prepare for CALPADS but not including “records transfer or any of the other long-term functional CSIS goals, which are also the goals of CALPADS.”⁴⁹ As noted earlier, funding supported by CDE and LAO for ongoing maintenance and quality measures related to student identifiers was rejected. According to the feasibility study reports for CALPADS and CALTIDES, system development costs will largely be paid for with federal funds. There is little in this recitation to suggest that California is becoming significantly more willing to invest its own money to get information to help policymakers in deciding how best to invest the \$45 billion they spend annually on public education. This reluctance to invest raises questions in this writer’s mind about, among other things, whether the ongoing, long-term commitment to communication and training that appears critical to the successful implementation and utilization of complex data systems will be made and about whether funding for new data initiatives will be driven by available federal dollars rather than a realistic assessment of the costs of effective development and implementation.

Access to the data. Stakeholders need to be able to access and use education data if the new California initiatives are to fulfill their potential. While current activities like DataQuest, Ed-Data, and SchoolMatters do make a fair amount of information available, often in user-friendly fashion for those interested in descriptive data on particular schools, districts, or counties, the more far-reaching benefits for policymakers will come from researchers (both inside and outside government) who undertake analyses aimed at determining which educational programs and what kinds of resource use really make a difference in improving student achievement and in narrowing achievement gaps.

States take different approaches to fostering the use of their education databases for analytical purposes. In Florida, for example, the state investment in the Education Data Warehouse has put data in a form suitable for research use; and staff in the Department of Education and the legislature’s Office of Program Policy Analysis and Government Accountability undertake many analyses themselves. Outside researchers can apply to FLDoE for access to data from the warehouse and must agree to abide by a number of conditions relating to security, training on the structure and content of the warehouse, and

⁴⁸ CSIS Strategic Plan, January 2006, p. 14, available at <http://www.csis.k12.ca.us/library/CSIS-Strategic-Plan-version-2%200-20060109.pdf>.

⁴⁹ <http://www.assembly.ca.gov/acs/committee/c2/hearing/may%2022%202006%20part%20ii%20public.doc>.

other matters. Agency staff must be allowed to review research reports before they are released. In North Carolina, where cuts to the state department of agency have severely limited the capacity of staff to undertake all the tasks necessary to make data usable for research, a North Carolina Education Data Research Center has been established at Duke University to create a portal to the immense amount of education data that North Carolina collects. A Center Advisory Board of members from the sponsoring consortium (Duke, the University of North Carolina, and the NC Department of Public Instruction) is responsible for setting access policies and reviewing applications to use the data. In Texas, researchers have applied more informally for access to education data through the Texas Education Agency (TEA), but a new law (HB 1, passed in June 2006) calls for the creation of up to three Education Research Centers at TEA, the Texas Higher Education Coordinating Board, or colleges and universities “to conduct research for the benefit of education in this state.” How this provision will be implemented is currently being decided. In New York school finance analyses of interest to state policymakers has been carried out for a number of years under agreements between the state Department of Education and a university-based Education Finance Research Consortium. Researchers in the consortium have undertaken the work of putting together the data files needed to conduct consortium analyses.

The analytical potential of new data systems is most likely to be realized under conditions in which funding is available to develop research-friendly data files and outside researchers as well as state employees can have reasonable access to them. Where state agencies themselves control access to the data (as in Florida and is planned in California), outside researchers may have difficulty if state employees are reluctant to give up data they’ve worked hard to develop (and want to use themselves) or if politics influences what data projects are approved. (Interviewees indicated that both issues have been known to arise in Florida). California might consider how outside, disinterested stakeholders could be involved in setting and carrying out policies relating to data access, especially for outside researchers, to ensure that appropriate requests are granted. If California does not develop a data warehouse or other means to prepare education data so that it can be used for research and analysis, it might consider sponsoring (and perhaps contributing funding to) an outside agency to do so.

Because this paper is being written as part of a project on school finance and resource allocation, a few final comments about finance data are in order. It was noted earlier that as of 2003-2004 all school districts are required to report financial data to the state using a common accounting structure. It was further noted that the SACS includes subcodes for school-level financial reporting, although this is voluntary; and common data definitions for school-level elements have not been developed. It is reasonable to ask whether California needs school-level

financial data to answer important questions about how resources spent relate to student performance and to identify resource gaps that may be related to achievement gaps.

The effort required to develop school-level data does not at this time seem justified, despite well-known limitations with finance data that can only be disaggregated to the district level. A host of problems have been identified in developing school-level finance data reporting systems, the most significant of which is the immense burden such systems place on local administrators.⁵⁰ Even if California did not already have enough new data initiatives on its plate, serious questions would exist about whether the benefits of attempting to obtain school-level data come anywhere near their costs.

A few years ago, the Joint Legislative Accountability and Review Committee (JLARC) in Washington State considered but rejected the idea of collecting school-level expenditure data, concluding that expenditures on staff (which account for a very large percentage of school and district spending) were far easier to obtain and could serve as a useful proxy for school-level expenditure data.⁵¹ A 2005 JLARC noted the potential advantages of school-level financial data in improving accountability, transparency, linkages with performance, and equity and observed that *districts* might find it advantageous to track expenditures to the school level in their own reporting systems.⁵² The costs and administrative complexities of attempting to develop consistent, comparable school-level financial reports on a statewide basis, however, remain daunting. Instead, California should concentrate on developing a reliable and detailed teacher

⁵⁰ Lawrence O. Picus, *In Search of More Productive Schools: A Guide to Resource Allocation in Education*, ERIC Clearinghouse on Educational Management, College of Education, University of Oregon, 2001.

⁵¹ State of Washington, Joint Legislative Audit and Review Committee, *K-12 Finance and Student Performance Study*, Report 99-9, September 15, 1999.

⁵² State of Washington, Joint Legislative Audit and Review Committee, *K-12 School Spending and Performance Review*, Preliminary Report, September 14, 2005.

(i.e., certificated staff) data system, including information on salaries and benefits, that can be linked to schools, students, and programs. Such a system could support analyses of many of the central questions about resource allocation without requiring whole new school-based financial reports.