

# POLITICAL ROADBLOCKS TO QUALITY DATA: THE CASE OF CALIFORNIA

---

RISHAWN BIDDLE

*RiShawn Biddle is a journalist and editor of Dropout Nation*

**I**n 1997, policymakers in California started building a comprehensive, longitudinal data system that would provide quality education data to politicians, parents, administrators, researchers and activists. Eleven years — and numerous laws, policy statements and blue-ribbon reports — later, the work remains incomplete. California’s difficulties in developing its data system exemplify the problems faced by other states in pulling together the student-, teacher-, school- and district-level performance data needed to conduct high-quality research and make wise decisions.

The state legislature established California School Information Services (CSIS) in 1997 to serve as a statewide repository for student data.<sup>1</sup> The CSIS system was designed to enable school districts to transfer individual student records electronically to the state using its State Reporting and Record Transfer System (SRRTS) instead of sending reports based on aggregate data. CSIS assigns unique student identifiers and collects basic information such as gender, school lunch (or socioeconomic) status, and grade and course completion information for individual students (with names and other personally identifiable information stripped out).

However, CSIS isn’t a truly comprehensive statewide system. Why? State officials never committed to fully funding its rollout. As a result, just 263 of

the state's 1,058 school districts—representing 60 percent of the state's public school population—send data to CSIS. And the records that are submitted to CSIS lack some desirable features: student-level records for English language learners (ELL) do not include information about the program or setting in which an ELL student is being taught, for instance. The agency doesn't collect or store test score data from the array of exams given by the state every year. Finally, it only stores data for seven years.

The rest of California's education data system—from its array of data collections to the test-score information collected by its vendors—isn't any better. Russlynn Ali, executive director of the California division of the Education Trust, the Washington, D.C. school reform group, declared in a 2007 brief on school data, "California's education data system barely merits the name: It is a confusing assembly of collection vehicles, aggregated at different levels, reported at different times, housed in a multitude of different databases and only linked manually according to the ever-increasing demands of federal and state reporting."

In March 2008, the Committee on Education Excellence, a panel convened by the state's governor, Arnold Schwarzenegger, pointed out that the problem isn't a lack of data. What is lacking, the committee noted, is a systematic effort to "collect, integrate and maintain the array of information available."

In order to remedy the shortcomings of the CSIS system, in 2002 the state legislature voted to create a new state data system, the California Longitudinal Pupil Achievement Data System (CALPADS). The design for CALPADS was not approved until November 2007; it took five years to complete the process of putting together feasibility studies, getting approval from the state department of finance, putting together requests for proposals, and procuring the actual bid. The system is currently scheduled to go online by the beginning of the 2009–10 school year.

Unfortunately, even when fully realized, CALPADS will be inadequate in many ways. And the problems California has encountered as it has attempted to build CSIS, and now CALPADS, have arisen in other states attempting to develop comprehensive data systems. As this case study reveals, there are three main failings which can emerge as states attempt to build high-quality data systems:

1. **Data systems are too narrowly focused on meeting accountability rules, restricting the array of data included:** The systems are created

only to comply with No Child Left Behind and state accountability requirements. As a result, they don't contain the wide range of data that can be used by teachers and administrators to effectively shape curricula and instruction—especially for the very socioeconomic groups the accountability laws are supposed to help. And the data collection processes supplying information into the systems “silo” data in ways that are oriented towards monitoring compliance. It's difficult to reorganize the information so it can be used for research and decision-making.

2. **The systems aren't fully integrating state and district elementary and secondary school databases:** Seamlessly connecting state systems to district-level databases—which hold the underlying files and information—by standardizing technology and processes at all levels is crucial. This integration can also help expand data capacity at the district level. But states aren't devoting enough technical and financial effort towards this goal.
3. **A lack of cooperation among K-12 and postsecondary agencies and institutions limits the scope of the data:** The sprawl of K-12 and higher education agencies (each with their own systems, technologies, data sets, interpretations of student privacy laws, and procedures), the varying quality of data in each system, and their mutual unwillingness to concede ground to the other impedes efforts to tie them together into unified regimes. A lack of unified governance intensifies the disputes.

For policymakers, administrators, advocates, researchers, and parents in other states, the struggles faced by the Golden State in these three areas offer lessons about obstacles that must be overcome.

### **Failing #1: Narrow Focus, Narrow Purpose, Limited Use**

A first step in developing a data system is deciding what information should be collected and for what purpose. California's experience in the development of the CALPADS system offers some insight into how a narrow focus and purpose can limit a system's usefulness.

CALPADS will store student-specific data including school enrollment, socioeconomic status, whether the student is an English Language Learner,

discipline records, and scores on the battery of achievement tests given by the state every year. Information on whether the student graduated from high school, dropped out, or received a General Education Development or special education certificate will also be collected and stored in the system. Much of this is data which must be collected to comply with the No Child Left Behind Act.

Ideally, CALPADS would store an even wider array of student-specific data. Student course grades, for example, won't be stored in CALPADS—nor will SAT and other college-readiness test scores, individual student attendance records, or information on vocational and special education programs.

Essentially, declared Education Trust's Ali, "A wide gulf lies between what the new data sets should and could tell us and what they will actually have the capacity to do."

### **Conflicting needs, dueling priorities**

The narrowness of CALPADS' holdings reflects the conflict over whether data systems should focus on collecting and storing data needed for compliance with state and federal laws or whether they should include a wider array of data and organize it in a way that is useful for broader decision making and research.

Traditionally, state-level school data systems were developed as key data delivery and compliance points for the U.S. Department of Education and for state policymakers, not as sources of information for school- or district-level managers, much less teachers. This priority has only grown in the past two decades with the passage of No Child Left Behind and accountability measures at the state level. The very efforts to make schools accountable for student academic performance and to improve data quality have, ironically, focused state and local education officials more on compliance than ever, at least when it comes to data.

For those administrators charged with monitoring compliance, the primary need is for aggregate student achievement data and the same data disaggregated by NCLB-specified population groups, not longitudinal measures of the performance of individual students. But this traditional emphasis on compliance ignores the needs of other parties with expanding roles in the educational landscape. Although data-driven research and decision making is a fairly new concept in education, it has been widely embraced by policymakers,

district-level administrators, teachers, researchers, parents, and advocates. Their information needs and uses, however, (*see sidebar: Education Data Users and Uses*) differ greatly from those of compliance-driven administrators. While aggregate data offers them some of the information needed, they really want student-level data that can help expand their knowledge of the driving forces behind student performance.

It is not just the lack of longitudinal, student-level data that limits the usefulness of state data systems. The data that are collected tend to be gathered and stored in idiosyncratic ways, resulting in isolated silos of data that cannot “talk” to each other.

California’s education department has 125 different data collections.<sup>2</sup> These range from the California Basic Educational Data System (CBEDS), used to collect student, teacher and classified staff demographic data, and the Standardized Account Code Structure, used to collect school funding and budget data, to the Student National Origin Report (SNOR), which is used to count foreign-born students and their countries of origin. The number and range of California’s collections isn’t exactly unusual: the Colorado Department of Education, for example, has 142 different data collections, according to Jan Rose Petro, the agency’s director of data collections.

In California, these collections and their related databases are generally geared for internal use by state- and federal-level administrators for monitoring

### **Education Data Users and Uses**

#### **U.S. DEPARTMENT OF EDUCATION**

**Data Needed:** Disaggregated achievement results by subgroup; Adequate Yearly Progress (AYP) for each school and district; teacher qualifications; program expenditures; program enrollments

**Why:** Ensure No Child Left Behind (NCLB) compliance, analyze national school performance and inform improvement efforts

#### **STATE POLICYMAKERS AND EDUCATION AGENCIES**

**Data Needed:** Standardized state test scores; percentages of students achieving proficiency

**Why:** Establish and monitor compliance with state standards; align curricula with standards; recognize achievement; keep administrators and teachers accountable for performance; provide technical assistance to districts and schools; program design; inform school choice

compliance by schools and districts. More than 70 percent of data collected by the state's education department is tied to federal requirements.<sup>3</sup> A four-decade-long expansion of categorical (or specially targeted) programs in the state—programs aimed at everything from aiding disadvantaged students to purchasing computers for classrooms—has also expanded this compliance orientation. In California, more than 60 such programs accounted for 30 percent of all education spending during the 2006–07 fiscal year. Few of the programs are as large as the state's \$3.1 billion special education program or its \$1.7 billion class-size reduction initiative, but they all require monitoring.

The structures of the 125 data collections are dictated by specific federal and state legislation; their focus, naturally, is on compliance. As a result, the data system doesn't make data very accessible to parties involved in research and decision making.

Not only does this unwieldy process of gathering and storing data produce state data systems that are not very powerful, the state system hamstring school districts. Many districts are trying to expand their use of data from simple compliance to designing curricula, instruction, and school improvement plans, but the arrangements of their own data systems are unavoidably affected by decisions made at the state level. In order to comply with mandated reporting, for instance, a district will develop one database to track students in Title I programs, another to collect data on ELL program participants, and a third

#### DISTRICT-LEVEL ADMINISTRATORS

**Data Needed:** Percentages of students achieving proficiency by school and subgroup; aggregated and disaggregated longitudinal student achievement data

**Why:** Help parents and community focus on student achievement; provide technical assistance to schools; NCLB and state accountability compliance; curriculum decisions, research.

#### SCHOOL ADMINISTRATORS

**Data Needed:** Student performance by grade, program, teacher and population group; percentages of students achieving proficiency by grade, program, teacher and population group; disaggregated longitudinal student achievement records; attendance data; graduation rates; individual student performance records.

**Why:** Keep focus on student achievement; structure curricula and instruction to student needs; track down students struggling

[continued]

to track students in migrant education programs—even though, thanks to the state’s heavily Latino population and agricultural sector, a student may participate in all three programs. The district ends up with a data system useful for compliance, not for school improvement.

### **Finance dictating structure**

Further driving the development of data systems toward compliance instead of broader data use is the matter of cost. As control over school funding has shifted from districts to statehouses, so has the competition for cash. Data system development competes with other priorities such as programs for disadvantaged children, music instruction, and class-size reduction initiatives. Unlike those programs, there are few champions for data systems development save for school reform advocates and those administrators and policymakers embracing data-driven decision making.

“Data systems are long term. They benefit the student, but not that year,” said Stefanie Fricano, an analyst with the state Legislative Analyst’s Office, which serves as both an advisor to the legislature on fiscal issues and an advocate for expanding school data systems in order to bring transparency to the system. “That is always difficult for people when they are trying to decide what to do with money.”

academically; school community focus on student achievement; focus operations.

#### **TEACHERS**

**Data Needed:** Student test scores; student achievement by program and population group; percentages of student subgroups achieving proficiency; student test scores; longitudinal individual student achievement trends; attendance; student performance in prior and subsequent grades.

**Why:** Diagnostic information on students’ learning needs; help focus both teachers and students on achievement; focus staff use of time; track students in

need of assistance; assist, in classroom curriculum decisions; create additional assessment items

#### **STUDENTS AND PARENTS**

**Data Needed:** Grades on assignments and courses; test scores; individual longitudinal achievement record; diagnostic information on students’ learning needs; school and district performance data.

**Why:** Assist parents and students in choosing best schools and programs for needs; help focus both on student achievement; inform progress against proficiency standards.

Policymakers and administrators at the state level already struggle to justify investments in data systems. They have little incentive to structure those data systems to focus on anything other than compliance.

The development of CALPADS reflects how fiscal considerations further drive the development of data systems towards a compliance orientation. In 2002, a year after Congress passed No Child Left Behind, state officials concluded that CSIS's operations would not provide the information needed to comply with NCLB's reporting requirements. Nor would CSIS be able to meet the requirements of the Public Schools Accountability Act (PSAA), the state accountability law passed by the legislature two years earlier.

Early in 2002, legislators didn't even consider developing a more extensive system. They were too concerned about having to pick up costs borne by school districts and other local governments. Article XIII B of the state's constitution mandates that the state government must reimburse districts and other agencies of local government if they are required to comply with state demands that would otherwise be considered "unfunded mandates." Among the requirements that would be considered unfunded mandates are any new data collections or data elements required by the state as part of any education legislation. The state department of finance—which directly advises the governor on spending issues—and the legislature tend to

**RESEARCHERS AND ANALYSTS**

**Data Needed:** Percentages of students achieving proficiency by school and subgroup; aggregated and disaggregated longitudinal student-achievement data; student performance by grade, program, teacher and population group; percentages of students achieving proficiency by grade, program, teacher and population group; disaggregated longitudinal student-achievement records; attendance data; graduation rates; individual student performance records.

**Why:** Ability to track the results of curricula and standards over time; inform school choice community and business and industry; percentages of students and subgroups achieving proficiency; school report cards; help parents and community to focus on student achievement; provide assistance to needy schools

**Source:** *Table developed by Robert M. Pailach, Dixie Griffin Good, and Ari van der Ploeg. State Education Data Systems That Increase Learning and Improve Accountability. Learning Point Associates and North Central Regional Educational Laboratory, June 2004.*



err towards interpreting new reporting requirements as unfunded mandates, the costs of which the state must bear. So when Senate Bill 1453, which established CALPADS, was crafted and passed by state legislators, it was only authorized to store information need to comply with the requirements of No Child Left Behind and PSAA.<sup>4</sup> (PSAA data collection isn't considered an unfunded mandate because it involves data that schools are already required to collect in order to receive funding under previous federal and state laws.)

The compliance orientation was further emphasized by the state's process for approving the technical and financial parameters of the information technology system. At the time of CALPADS' development, the department of finance was charged with overseeing this process; that role has since been handed over to the state's chief information officer. The department of finance strictly interpreted the legislation that established those data systems to ensure that the data elements being included in them did not violate the unfunded mandates clause.

Tensions between the education department and the department of finance were exposed in a January 2005 review of a report which included design and technical specifications for CALPADS. In that review, the finance department concluded that the initial plan for the systems contained "data elements and/or collections" related to ELL students not specified either by law or for NCLB compliance. Education department officials explained to finance department officials that school districts were already required to collect those elements as part of the Language Census data collection, according to Keric Ashley, the education department's director of data systems. Eventually, finance officials conceded that point, approving the project.

An effort to expand the data stored in CALPADS came in 2005 in the form of Senate Bill 368, a bilingual education bill authored by State Sen. Martha Escutia. A provision, amended into the bill early on, would have required the creation of a separate database in CALPADS for tracking the performance of individual ELL students—including test scores, course completion information, and whether the students were eventually mainstreamed into regular classes—in a longitudinal manner. The associated cost of expanding CALPADS to include this database, however, led to the provision being stripped out of the bill upon its passage a year later.

## **The consequences of limited structures**

One consequence of designing a data system to focus only on compliance is illustrated by a report released in January 2006 by the American Institutes for Research (AIR) and WestEd, a San Francisco-based education think tank, on the instruction of ELL students.<sup>5</sup>

California has 1.6 million children in need of English language instruction—one in every four students. In 1998, voters approved Proposition 227, which required schools to instruct ELL students by immersing them in English rather than using bilingual instruction, but no one is sure whether immersion is actually working as intended.

AIR teamed up with WestEd on a five-year project commissioned by the state legislature to analyze whether English immersion instruction is better than bilingual instruction. Unfortunately, student-level data—especially about the instructional setting in which a student is being taught—isn't available statewide in California. Aggregate data on the performance of ELL students are distributed across at least four state data collections, each with their own collection periods. As a result, it is difficult to combine them.

All this limited the analysis that AIR and WestEd could perform. In their report, released in 2006, they wrote that they were unable to determine whether traditional bilingual instruction methods or full English immersion was more effective at improving the academic performance of ELL students.<sup>6</sup> AIR and WestEd researchers conceded “limitations in statewide data made it impossible to definitively resolve the longstanding debate.”

## **Failing #2: Failure to Integrate State and District Technologies**

The second critical component of comprehensive school data systems—especially at the K-12 level—is integration with the district-level systems and databases that initially collect and store the data. The key to this is standardizing the underlying technology of both systems in order for data to be easily transferred electronically.

For state-level administrators, integrating state and district systems allows data to be collected, stored, and accessed in real time, making for smoother, more accurate transfers of information. It can also help reduce the burden of paperwork faced by districts in meeting overlapping state and federal reporting requirements. Data system integration can also spur districts to

improve existing systems and improve the quality of the data stored at the district level.

Achieving such integration, however, requires policymakers and administrators at the state level to include districts as they design the system and its underlying processes, and also to provide financial and technical support for districts. California offers lessons in how not to do so.

### **Lofty goals, sluggish follow-through**

Early on, California's legislature recognized the importance of integrating state and district-level systems along with standardizing technology among districts. Through the law that created CSIS in 1997 and later legislation, the agency was charged with helping the districts develop "comparable, effective, and efficient pupil information systems" for their own operations and reporting to state and federal education agencies. Legislators wanted 90 percent of districts to submit data to CSIS in a standardized, electronic format by the 2004–05 school year and sought to encourage it in these ways:

- CSIS would oversee the implementation of statewide student identifiers to be used at the district level.
- Schools would be able to electronically transfer individual student transcripts, test score results, even health and discipline records to CSIS. This would lead to technical and data management standardization and integration between state and district systems.
- Aggregate data collections would gradually be replaced by reports generated from the individual student data and CSIS would work with the Education Department on streamlining the latter's 125 data collections. The two initially identified 40 aggregate data collections to be transitioned from traditional paper delivery to electronic transfer.
- Technical advice would be given to districts, especially when it came to sending data to CSIS.

Although CSIS has successfully transitioned districts into using statewide student identifiers, it hasn't made much headway in its other goals.

An electronic data transmission system was created, but by 2005–06, just 263 districts were using the system. Districts that participate submit

individual-level data to CSIS using the SRRTS software, and CSIS generates the necessary aggregate data reports and sends them to the California Department of Education. Districts that do not participate generate their own aggregate data reports and send them to the California Department of Education themselves. A lack of sustained funding for integrating CSIS with district-level systems is the main culprit behind the low level of participation, but another factor must surely be the fact that CSIS and the education department haven't succeeded in transitioning many data collections to the new system. By 2008, only five data collections were handled using individual-level data submitted to CSIS.

And submitting data electronically to CSIS is not simple. To prepare data for submission to the state, district-level administrators use a 214-page data dictionary to find the proper codes. They must comb through five different Microsoft Word files—some of the files as long as 54 pages—in order to learn the requirements needed for creating the files that will be sent through the system. A 62-page guide details how each file being transferred through CSIS must be put together for processing. Five Excel spreadsheets map out other data requirements. All of this work is required to submit individual-level data to CSIS to satisfy five of the state education department's data collections. For the state's other data collections, districts must deal with other manuals, file-creation rules, schedules, and formats.

### **Why the effort failed**

Why wasn't the legislature's mandate to integrate state- and district-level data systems ever fully realized in California?

The problem begins at the state level. Back in 2002, a report on the data processing and management practices of the state education department by MGT of America, a Tallahassee, Fla., consultancy, noted that the state education department was struggling with its key role in the state's data system:<sup>7</sup>

- Data collection within the department was highly decentralized; each program office had its own process for collecting, processing, and storing data.
- Coordination of data among those offices was minimal; essentially no one could get a full understanding of how data were managed within the department.

- Data management within program offices wasn't rigorous, aligned to any kind of data management standards. The department itself didn't have a common vision about how data should be processed, collected and stored.
- Data dictionaries weren't standardized throughout the department; thus no consistent system for naming and defining datasets and data elements throughout the department.
- Electronic data transmission barely existed; paper submission was heavily relied on.
- Data collections involved inconsistent units of analysis or inconsistent time periods.

Part of the problem lies with the penchant of California state legislators for using categorical programs to fund schools. The original goal behind creating categorical programs was to force specific reforms at the school district level and keep tabs on their progress. But as the number of categorical programs grew, so did the number of offices set up to monitor these programs. Each office and program developed its own data collection process. This contributed to a confusing sprawl of data collections and databases at both the district and state level. Although the education department has since moved to create a data oversight office charged with streamlining processes and standardizing data dictionaries, this office still struggles to serve both districts and other data decision making parties.

The main reason behind the lack of integration of state and district technologies, however, was lack of sustained funding. California's experience shows how states are struggling with the fiscal price of their expanded role in funding and structuring education policy.

Since the 1970s, when property tax revolts and lawsuits over equal funding of poor and wealthy schools began to reshape the public education landscape and move power away from school districts, state governments have become the primary arena for education policymaking. California's experience is all too familiar on this front. Beginning in 1971, when the legislature — heeding the call from homeowners about rising property taxes — enacted “revenue limits” or caps on income districts could generate from property taxes, the state has become the dominant player in deciding

education spending at the local level. That role grew in 1978, when voters first passed Proposition 13, which essentially froze and then reduced property tax revenues for school districts. A decade later, voters passed Proposition 98, which required that a minimum percentage of the state budget be spent on education.

The result was that the state's share of education spending in California increased, from 34 percent in 1972 to 67 percent by 2005.<sup>8</sup> (The share of spending paid for by local revenues declined from 60 percent to 22 percent in that period.) Nationwide, the average share of education spending by state governments grew from 38 percent in 1972 to 46 percent by 2005, according to the U.S. Department of Education.

The increased burden on the state was intensified in California by Article XIII B of the state constitution, which mandates that the state government must reimburse districts for complying with reporting requirements that otherwise would be considered "unfunded mandates." As a result, legislators, governors and the state department of finance look for ways to limit state-level costs when developing data systems, which ultimately limits the integration of state- and district-level systems.

Early on in the development of CSIS in 1997, legislators debated whether to make district-level participation mandatory. The ultimate deciding factor was the cost. In order to avoid imposing any "unfunded mandates," legislators made participation voluntary; districts could decide whether they wanted to submit individual-level data to CSIS. In exchange for participating, districts would receive one-time implementation grants covering 50 percent of a district's cost of implementation.

But by 2001, funding voluntary participation became a challenge. That year, officials overseeing CSIS proposed to spend \$23 million on implementation grants, but the legislature only allotted \$11 million, financing implementation grants for a mere 98 districts. That same year, the legislature attempted to guarantee 90 percent district participation by the 2004–05 school year by passing AB 295, which would have required the state to spend \$104 million over four years to reach that goal. But Governor Gray Davis vetoed that bill, arguing that the state would likely have to cover costs above the \$104 million because of the unfunded mandates clause. Three years later, during one of the state's periodic budget crisis, the legislature cut out implementation grants altogether, stalling the expansion of the program to other districts.

Meanwhile CSIS began to find that, if anything, districts needed additional help in understanding the technology requirements for integrating their systems with the state system. Many districts had datasets of extremely poor quality that would have required significant cleanup before they could participate. A lack of adequate staffing and training to run existing systems also made it difficult for districts to take steps towards working with the state on systems integration.

CSIS's own mission of getting districts up to speed is itself compromised by low staffing. Of the agency's 53 employees, just 11 work on assisting districts with their data processing issues and questions. This lack of manpower limits the help districts can get for their data processing needs.

The launch of CALPADS in 2002 shifted the focus away from expanding CSIS. It also marked a move toward mandatory participation by districts in the state data system. In establishing CALPADS, legislators argued that in order to meet the accountability rules contained in both No Child Left Behind and the PSAA, all school districts would need to integrate their systems with that of the state. Any school district accepting Title I funds and state general purpose funding (or base operational funds) doled out on the basis of enrollment—essentially every school district in the state—had to go along. “By taking federal funding, they are making a commitment to reporting anything the federal government is funding,” said Ashley, the state official overseeing CALPADS.

Having decided that district-level participation in CALPADS would be mandatory, state policymakers needed to develop a strategy to get those districts not participating in CSIS's individual-level data collection up to speed technologically so that they could be integrated into the new system. Some 300 school districts have enrollments of 300 or fewer students, and the quality of their data systems is mixed. Some districts are storing data using Excel spreadsheets and FileMaker software, with a secretary or another staffer handling data processing needs. Integrating these districts into the state system will be an arduous task for the districts and the education department alike.

The state opted to begin the transition to CALPADS in the 2006–07 school-year by funding a program called Best Practices. Under the program, school districts with enrollments less than 1,800 that implemented the unique student identifier (and weren't already participating in the CSIS data collection) would

get funds to help build out new data systems that could be easily integrated with the new state system. As part of the process, districts would clean up student standardized test files and improve data management practices so that the districts could begin delivering data electronically.

Funding for Best Practices, however, was contentious from the start. The Legislative Analyst's Office, though supportive of the program, recommended that legislators trim the original \$30 million proposal by half. The department of finance opposed the program, arguing that implementation grants weren't needed until CALPADS was up and running, according to Janet Hansen, a senior policy researcher with the Rand Corp. After some wrangling, Best Practices was funded to the tune of \$31 million, to be spent from 2006–07 through 2008–09.

Attempts to increase funding for Best Practices ran into roadblocks. A plan to boost funding for Best Practices by \$65 million (along with an extension of the program into the 2009-10 school year) was scotched before its final passage. Governor Arnold Schwarzenegger included some funding for Best Practices in his proposed 2008–09 budget; but an impasse between state legislators and the governor over the overall budget may eventually mean that the program will no longer be funded—just as CALPADS prepares to go online.

### **Failing #3: Failing to Unify K-12 and Postsecondary Data Systems**

Since 1994, 38 states have formed P-20 councils of some kind to increase the alignment of their preschool, K-12, and higher education systems, according to *Education Week*, in its most recent “Diplomas Count” report. But achieving the goal requires unifying elementary-secondary and postsecondary data systems, which currently operate independently of each other. California's experience offers a sober lesson in how educational governance structures and turf battles can frustrate such unification.

As the state embarked on developing a comprehensive, longitudinal data system at the K-12 level in 1999, it also began moving towards integrating its multiple systems at the postsecondary level. That year, the legislature reorganized its higher education oversight body, the Postsecondary Education Commission (CPEC), and charged it with connecting the data systems of the University of California (UC), California State (Cal State), and the state's community college systems. This new database was to conform to the one being developed by CSIS for K-12, creating the potential for unification. In



its database, CPEC was supposed to collect student transcripts—including information on course completion, grades, unit hours earned, and degree-seeking status—along with student-level socioeconomic data.

From the get-go, CPEC struggled to get the universities on the same page. The community college system had been supplying student-level data to the commission since 1993, long before CPEC was reorganized and charged with unifying higher-ed data systems. The data include a student's high school of origin, degree-seeking status, and grade point average. The UC and Cal State systems, on the other hand, were more reluctant to release data because of their interpretations of the federal Family Education Rights and Privacy Act (FERPA) and the state's own array of student privacy laws. Only in 2005—six years after the legislature charged the commission with its task—did CPEC begin collecting data from them. So far, files collected from the central offices of the University of California and California State systems don't contain any course completion data at all because such information is located in files on university campuses and isn't transferred to either system's central database. They do contain such student-specific information as scores on SAT and ACT exams, credit hours earned, and degree-seeking status.

By law, CPEC and the university systems are supposed to meet regularly to advance the integration of data systems and develop a common data set that includes socioeconomic and course information. This isn't happening. University officials are unwilling to work with the agency because, they say, CPEC fails to consult them about how the data it receives will be used in its own research projects; the commission, for its part, notes that universities do get to review research before it is published. The lack of progress on this front has done little to improve CPEC's already low reputation among state legislators. "Nobody trusts their opinion anymore," said Amy Supinger, a consultant to the state senate's Budget and Fiscal Review Committee. CPEC has approached the Association of Independent California Colleges and Universities about accessing the data of its members, but no progress has been made on that front.

While CPEC struggled to integrate university data systems, the legislature took another step towards P-20 data system unification in 2003 by funding the California Partnership for Achieving Student Success (Cal-PASS), an Encinitas-based nonprofit group, to help link university and district-level data systems

and develop longitudinal tracking of student performance. In its own data system, Cal-PASS collects at least five years of longitudinal data from 4,000 participating K-12 schools and colleges. (Participation is voluntary.) Like CPEC, Cal-PASS stores student-specific information on degree completion status and high school of origin, but it also has access to transcripts and course grades not contained in CPEC's collection.

Even if the university systems were more cooperative, a major technical barrier remains: a lack of a unique student identifier used by all educational institutions. At the K-12 level, a unique identifier has been used for tracking data since the 2004-05 school year; high school seniors are now jotting down those identifiers on UC and Cal State applications so that the schools can access the students' records through CSIS and eventually, CALPADS. Colleges and universities, however, haven't adopted the K-12 identifier or developed a uniform system of their own. Within UC, each campus issues its own identifier; student movement is not tracked within or outside the system. A student transferring from, say, the University of California, Los Angeles to UC Santa Barbara is issued a new identifier upon admittance.

### **Governance structures that impede data unification**

At the heart of California's problems are governance structures that impede cooperation on data systems unification. At the K-12 level, governance is divided between the state board of education and a secretary of education—both appointed by the governor—and the state education department (controlled by an elected superintendent). There is also the Fiscal Crisis Management and Assistance Team (FCMAT), which manages CSIS and handles fiscal affairs within the education system, and the Commission on Teacher Credentialing, the teacher certification agency.

Governance of the university systems is even more unwieldy. Although CPEC oversees the higher education system, the UC, Cal State, and community college systems function independently, each with their own systems, procedures, and data sets. Even within institutions, governance is complex. Although a chancellor oversees the community college system, each college also reports to a regional board. Each campus in the UC system has an academic senate that shares power with campus-level administration.

With so many institutions and a lack of an overall governing body, it is difficult to get all the parties at the table. The result is predictable: little gets

done on P-20 unification. “Data and information systems are one of the victims of the state’s current convoluted governance structure,” according to Governor Schwarzenegger’s Committee on Education Excellence. And there has been little recent effort to push for a change in the status quo. In 2004, the state schools superintendent, Jack O’Connell, announced with great fanfare the formation of a 64-member P-16 council in order to build consensus among all stakeholders on unifying the education system, including integrating data systems. Three years later, the council has issued reports on reforming high schools. But so far, data systems integration hasn’t been on its agenda.

In November 2007, Governor Schwarzenegger’s Committee on Educational Excellence recommended the creation of a commission to take over all current data systems and create a new one that unifies not only data systems at the state level, but those at the local level that often don’t match up technologically. The governor, however, ignored that recommendation; instead, he proposed in his state of the state address to create an education data commission to develop additional recommendations. That investigative body has yet to be formed.

In June 2008, Senate Budget and Fiscal Review Committee Chairman Denise Moreno Ducheny proposed to eliminate CPEC by the 2010–11 fiscal year and hand over its data management function to the state library. A lack of a plan for handling CPEC’s other functions, along with lobbying by members of the agency’s governing board, quashed that effort.

### **Steps Toward More Comprehensive Data Systems: Two Approaches**

California’s experience offers lessons to policymakers in other states on how not to proceed with developing comprehensive data systems. Florida has taken a very different approach.

States have different traditions when it comes to developing a “culture of data” in which data-driven decision making is encouraged and policymakers focus on improving data systems at all levels. Only a few have a strong tradition of supporting data system capacity at the district level. California has always shown “lukewarm support for education data system development” at all levels, according to Rand’s Hansen in a report on the development of the state’s data systems released last year.<sup>9</sup>

This contrasts with Florida. Since 1970, policymakers in the state have taken an active, involved approach to encouraging districts to improve data systems; they have also reduced reporting burdens, streamlined data reporting,

and helped districts improve their ability to use data in decision making. Beginning in 1976, data sets and data elements were standardized for all educational agencies and institutions. By 1985, state- and district-level systems were integrated through the Florida Information Resource Network (FIRN), which served as the backbone of the state's current school data system; by 1991, districts could transfer student records to one another through the FIRN, encouraging data sharing among districts (and also, with universities).

As a result of these and other moves, the Sunshine State is one of just four states cited by the Data Quality Campaign as positioned to have all ten basic elements of a comprehensive, longitudinal data system in place by the end of this past school year. California has only six of the ten elements in place.

The two states have faced the key challenges of creating a statewide data system in very different ways.

1. **Taking a broader view of data:** While Florida's data system is designed to help districts and the state comply with federal and state regulations, it is also becoming more useful for all parties. Teachers will soon be able to access student-specific data on the Sunshine Connections portal and use tools that will help with designing instructional efforts. The development of a data warehouse, in which student-level data is stored along with information from other state agencies and institutions, also allows for researchers to conduct a wide range of longitudinal research.
2. **Incorporating districts in data system design:** As noted earlier, Florida has tailored its system so that all sides gain; the state can get the information it needs while the reporting burdens of districts are reduced (and districts get a wider range of data). In 1987, the state began replacing aggregate data collections with individualized student- and teacher-level data reporting in FIRN; this simplified district-level reporting while moving the more tedious job of aggregating data and generating reports to the state level.
3. **Requiring the entire education sector to cooperate on data system integration:** Cutting through complex educational governance structures is critical to integrating K-12 and postsecondary systems. Policymakers in Florida have found a way to make this happen. Leadership from governors as diverse as Lawton Chiles and Jeb Bush helped universities

overcome their reluctance to share data. And universities in Florida have a lot to gain. Since education databases there have been linked to other state databases containing information about employment, universities can assess their own performance by tracking how graduates perform in the workforce after leaving college.

Many of the difficulties faced by California in attempting to build a comprehensive data system involve state-specific challenges. As Nancy Smith notes in her paper in this volume, the cultural norm in California is that the state department of education does not collect data without a specific mandate and funding. And the part of the state constitution prohibiting “unfunded mandates,” has meant that the California Department of Education must reimburse districts for the effort involved in submitting any data that are not strictly required in order to comply with state or federal law. Compounding this problem is a state department of finance and state legislature that are very aggressive about stopping the state from imposing costs on districts. This makes it extremely expensive for the state to collect the data it needs from districts, even though that very same data would be useful to districts.

California was also hampered by its propensity to fund its schools via a large number of different categorical programs, each with its own data requirements, which may have fostered the tendency to organize the data into silos.

Finally, there was a real lack of leadership behind California’s efforts to build a statewide education database. Without the governor or powerful legislators taking this project on and seeing it through, and without the state superintendent or state board of education making it a high (and sustained) priority, it was impossible to cut through the many fiefdoms with competing interests and narrow focuses to make anything big happen.

But in many ways, California is not a special case. The tendency to gather data in many separate collections and to store data in databases that don’t connect with one another is common. The tendency to only collect the data strictly required for federal and state compliance is also common. The difficulty of financing data systems is typical throughout states without strong cultures of data-driven decision making. And the inability to get higher-ed institutions on board with sharing data for a statewide database is something nearly all states have experienced.

## **Glossary**

### **CALPADS**

California Longitudinal Pupil Achievement Data System. Launched by the state in 2002 and expected to be operational in the 2009–10 school year, it will collect such individual student-specific data as socioeconomic status, discipline records, and scores on state assessments.

### **Cal-PASS**

California Partnership for Achieving Student Success. A partnership of K-12 and higher education institutions authorized by the state legislature to foster linkages between K-12 and higher education data systems on a voluntary basis.

### **Cal State**

California State University System

### **CALTIDES**

California Longitudinal Teacher Integrated Data Education System. This database will include a unique identifier for each teacher, credentials for each subject taught, and how the credential was achieved.

### **CBEDS**

California Basic Educational Data System. The California Department of Education's collection of aggregate student and staff demographic information.

### **CPEC**

California Postsecondary Education Commission. The state higher education oversight and coordination agency. It is tasked with unifying the data systems of the state's three university and college systems.

### **CSIS**

California School Information Services. It oversees the implementation of the unique student identifier (SSID) and operates the State Reporting and Record Transfer System.

**ELL**

English language learner. Students learning English as a second language.

**FCMAT**

Fiscal Crisis Management and Assistance Team. Run by the Kern County Office of Education, it operates California School Information Services (CSIS).

**FERPA**

Family Education Rights and Privacy Act. A federal law that limits access to individual student data to certain parties.

**FIRN**

Florida Information Resource Network. The initial effort by Florida's state government to develop a fully longitudinal data system.

**PSAA**

Public School Accountability Act. The state's standards and accountability law, which created the Academic Performance Index, a school performance measurement system similar to the No Child Left Behind Act's Adequate Yearly Progress measurement.

**SNOR**

Student National Origin Report. One of the California Department of Education's data collections.

**SRRTS**

State Reporting and Record Transfer System. Operated by California School Information Services, it allows school districts to transfer individual-level data that can be used to generate reports for five data collections (including CBEDS) to the state Department of Education.

**UC**

University of California System

## **Endnotes**

- 1 CSIS is an agency operated by the Fiscal Crisis and Management Team (FCMAT)—a quasi-state operation—and overseen by the state education department. A glossary of acronyms and other terms can be found at the conclusion of this chapter.
- 2 Hansen, Janet. “Education Data in California: Availability and Transparency.” The Rand Corp., November 2006.
- 3 Edwards, Sonya. “Reducing Data Collection and Reporting Burden on LEAs.” CDE Education Data News, Issue 2. California Department of Education, Spring 2007.
- 4 California received a grant of \$3.2 million to assist with the development of CALPADS from the U.S. Department of Education’s program supporting longitudinal data systems.
- 5 Parrish, Thomas B., Maria Perez, Amy Merickel, Robert Linqunti, et. al, “Effects of the Implementation of Proposition 227 on the Education of English Learners, K-12: Findings from a Five-Year Evaluation.” American Institutes for Research and WestEd, January 2006.
- 6 Parrish, Thomas B., Maria Perez, Amy Merickel, Robert Linqunti, et. al, “Effects of the Implementation of Proposition 227 on the Education of English Learners, K-12: Findings from a Five-Year Evaluation.”
- 7 “Data Management Study of the California Department of Education.” MGT of America, April 2002.
- 8 Timar, Thomas. “How California Funds K-12 Education.” Institute for Research on Education Policy and Practice, Stanford University, September 2006.
- 9 Hansen, Janet. “Education Data in California: Availability and Transparency.” The Rand Corp., November 2006.