

A Hunger for Information: California's Options to Meet its Statewide Education Data Needs

Part four in the series:

California Education Policy, Student Data, and the Quest to Improve Student Progress

By Colleen Moore and Kathy Reeves Bracco

California's education data, at least as they currently are collected, managed, and made available, are not sufficient for understanding and addressing the needs of California's students. That's the conclusion of the first three reports in the series [*California Education Policy, Student Data, and the Quest to Improve Student Progress*](#), which seeks to shed light on the role and potential of student-level data systems in California (see About This Series). This fourth report finds that California lags behind many states that have already developed valuable ways to use cross-sector statewide longitudinal data systems (SLDS) to improve policy and practice and to provide critical information to policymakers, educators, taxpayers, and students and their families. Fortunately, California has a range of options for developing an SLDS; there are numerous examples in other states that illustrate the possibilities. These data systems use a variety of structures and processes to link and manage student data across education systems to gain valuable information for improving policy and practice.

About This Series

California Education Policy, Student Data, and the Quest to Improve Student Progress

This brief is the last in a four-part series examining California's approach to gathering and sharing longitudinal data about students' progress through the state's education systems:

- [*Gaps in Perspective: Who Should Be Responsible for Tracking Student Progress across Education Institutions?*](#) An analysis of the perspectives of state and local leaders on who should be responsible for gathering and sharing data about students' progress.¹
- [*California's Maze of Student Information: Education Data Systems Leave Critical Questions Unanswered.*](#) An overview of student-level data collected and maintained in California, a summary of past efforts to develop a more comprehensive system, and an exploration of the strengths and weaknesses of the state's approach to education data.²
- [*Scaling Goodwill: The Challenges of Implementing Robust Education Data Sharing through Regional Partnerships.*](#) An analysis of some local and regional efforts to share data across institutions and systems that includes the benefits and challenges of participating in these efforts.³
- *A Hunger for Information: California's Options to Meet its Statewide Education Data Needs.* An exploration of lessons for California from other states' efforts to improve their education data systems, and some conclusions concerning a path forward to improve California's data systems for use in understanding and improving education policy and practice.

In this report, we briefly summarize the main findings from our previous reports, as they lay the groundwork for understanding why California’s education data systems are inadequate. We then examine the experiences of other states that have developed SLDSs, in order to understand the benefits and challenges of these systems and to clarify California’s options for developing its own statewide data system. In the report’s final sections, we offer an analysis of policy alternatives for California and recommend that policymakers create a state data agency, or an office within an existing state agency, that is tasked with developing an SLDS, creating standard reports and data dashboards for various audiences, and managing access to the data for research to inform education policy and practice. This report provides a detailed description of the issues and options related to developing an SLDS in California, as a resource for those who might be involved in deliberations about that possibility. For a short summary of our research findings and recommendations across this series, see [the research summary on our website](#).

What is an SLDS?

We define a statewide longitudinal data system, or SLDS, as a data system that connects student records across two or more core agencies from among early learning, K-12, postsecondary education, and workforce.⁴

California’s Education Data Systems Are Inadequate

Throughout this series, we asked stakeholders about the status of education data in California and whether creating a statewide cross-sector student data system is a necessary or feasible goal. We found a divergence in views about California’s approach to tracking, sharing, and using longitudinal data about student progress. State policy staff had heard calls for a statewide data system, but were skeptical of its utility and apprehensive about its costs. They thought local/regional data-sharing efforts were a good alternative. Local education leaders, on the other hand, emphasized the importance of having access to data for tracking students across the K-12 and postsecondary education systems. They were concerned about the challenges and costs associated with creating those systems locally and suggested the state take the lead in setting up a data system statewide.

“There’s a hunger for information that goes beyond counting, that actually takes a more holistic view about student outcomes, and that requires these advanced data linkages with longitudinal data.”

– John Armstrong, senior policy analyst, State Higher Education Executive Officers Association

Each of the systemwide offices of the state’s four systems of public education—the California Department of Education (CDE), the California Community Colleges (CCC) Chancellor’s Office, the California State University (CSU) Chancellor’s Office, and the University of California (UC) Office of the President—has a rich student data system. However, the data are disconnected and generally inaccessible for addressing the information needs of policymakers and education leaders. Despite fairly good data in each education system, the inability to link these data on a statewide basis leaves the state unable to answer many critical questions (see *Some Critical Questions California Cannot Answer*), particularly about student outcomes across education systems. Being able to answer these kinds of questions is crucial to designing and implementing effective education policies and to engaging in continuous program improvement at schools and colleges across California.



Some Critical Questions California Cannot Answer

The current structure of California's data systems leaves the state unable to answer numerous questions that cross the boundaries of the K-12 and postsecondary systems and into the workforce. There are likely significant equity implications related to each question that can only be understood with disaggregated data across sectors. Examples include:

- Which high school graduates from which schools are prepared to succeed in college?
- Are districts that receive extra funding through the Local Control Funding Formula increasing the proportion of their students who enroll in college, ready to succeed?
- What are the workforce outcomes of high school graduates who do not go directly to college? How many enroll in college later, and what are their outcomes?
- What happens to students who drop out of high school? Or drop out of a postsecondary institution?
- How many high school graduates meet eligibility requirements for admission to CSU/UC, but do not apply, and what happens to them? What happens to eligible students who do apply to CSU/UC but are not admitted due to capacity constraints?
- What are the outcomes of CSU/UC teacher candidates, and what is their readiness and rate of retention when they become employed by the state's K-12 schools?
- How many students take courses in the CCC while simultaneously enrolled in CSU/UC, and what is the impact of such course-taking on their graduation rates and time to degree?
- How successful, by major and degree/credential, are CCC/CSU/UC graduates in the workforce?
- How do all of these outcomes (and others) vary by students' race/ethnicity, income level, region of the state, or other important factors?

Since some state policy leaders suggest that this problem could best be addressed at a local/regional rather than at a statewide level, we examined several regional data-sharing efforts. While there are a number of regional efforts underway to share data, most are in early planning stages or are sharing only limited or aggregate data. These data efforts face significant challenges and, while valuable as mechanisms for collaboration, they are not an efficient or effective substitute for a comprehensive statewide data system. Even when successful at sharing student-level records, such efforts provide incomplete coverage, both statewide (most institutions do not participate in these data-sharing agreements) and within regions (some students move out of their regions for postsecondary education). In addition, these regional approaches can be difficult and costly to manage and sustain, especially for smaller school districts with limited resources or insufficient capacity for data sharing.

Research results from the first three reports in this series indicate that California's current education data systems are inadequate. They do not meet the needs of policymakers and educators for information that spans students' experiences across education systems. The remainder of this report describes our findings from additional research that examines the experiences of other states with statewide longitudinal data systems and the possible implications for California.

Experiences in Other States Offer Lessons for California

The Education Commission of the States reports that, as of November 2016, 37 states have an SLDS that links records for at least two of the four core state agencies (early learning, K-12, postsecondary education, and workforce), and 16 states have a full P-20W (preschool through workforce) statewide data system.⁵ The number of states creating these data systems has increased dramatically over the last decade.⁶ To understand the landscape across the country, we interviewed national experts from four organizations that study SLDSs, as well as individuals in six states who participate in or oversee the SLDSs in their home states.⁷

The growth in SLDSs has been driven, in part, by states' desire to improve education outcomes by using data to drive changes in policy and practice, according to interviewees. Growth has been facilitated by the federal government through funding from No Child Left Behind, Race to the Top, and SLDS grants.⁸ Others who have been supporting these efforts to establish or improve SLDSs include state governors and legislators concerned about improving the workforce and economy, and organizations such as the Data Quality Campaign.⁹ Regardless of the governmental or institutional body supporting the development of these data systems, our interviewees said the main reason that states have implemented the systems is the need for better information about what seems to be working (or not) to improve student progress along the educational pipeline.

“What’s driving [the development of data systems] is a recognition of how impactful data use can really be in improving college access and success, especially for low-income students, students of color, and other underserved populations. There’s a recognition of the value of data in informing decision making, both at the practitioner level and at the policymaker level.”

– *Mamie Voight, vice president of policy research, Institute for Higher Education Policy*

A Variety of Options for SLDSs

While each state approaches its SLDS slightly differently, there are two key factors that states must consider:

- Data system governance: Where are the data housed? This also determines who manages and coordinates the data.
- Data system structure: What is the best model or structure for linking the data?

Decisions about where to house the data may depend on the culture of a given state. Some states with a higher education coordinating board have found that this entity is a logical place to coordinate shared data. Other states have chosen their department of education, a state agency unrelated to education, or, in some cases, a university or other third party organization to house and manage the data. Regardless of where the data are housed, states also need to establish policies regarding the purposes for which the data can be used, who will have access to the data, and what role the managing entity serves, if any, with regard to research and dissemination.¹⁰ (See sidebars throughout this section for descriptions of how several states approach these issues).

Cross-sector student data in **Washington** are housed in a centralized data warehouse maintained by the Education Research and Data Center (ERDC), part of the forecasting division of the Office of Financial Management (the governor's budget agency). The effort began with a recommendation from a blue ribbon commission and is now established in a statute that requires data-sharing agreements between the ERDC and the state's education and workforce agencies. A federal SLDS grant provided support for establishing the governance structure, hiring researchers/analysts, and building the data warehouse. Governance includes three committees (data stewards, data custodians, and research and reporting) composed of ERDC staff and representatives from the participating agencies. The ERDC answers queries and conducts research only for issues that require cross-sector data.

In relation to the structure of the SLDS, some states have built a “centralized” system, where data from all agencies are collected in a data warehouse. A second model, at the other end of the spectrum, is a “federated” system, in which each participating agency continues to house its own data, but custom data sets are created by drawing data from each agency as needed to address particular research purposes. Each of these models has benefits and drawbacks. Centralized data warehouses are built on proven technology and allow for faster access to data for analyses, as the data are already matched and held in a single system. However, the data are only as current as the most recent upload. Federated systems may avoid turf battles over the location and control of a centralized data system. But they involve more cumbersome processes for the agencies to link the data for every individual use and require the development and maintenance of multiple data-sharing policies.¹¹ There are also hybrid models that incorporate elements of these two models. For example, a state might have a centralized warehouse containing student information from all of its education sectors, but use data-sharing agreements with other state agencies to match workforce, social services, health, or other information as needed.

We asked interviewees to reflect on how their states determined the structure and governance of their SLDSs. Several interviewees said that decisions about where to house the data depend to a large extent on the entity's capacity to manage data and do research. A common concern for institutions in agreeing to have their data housed elsewhere, in a centralized warehouse, is how their data will be shared, a concern often addressed through strict protocols for approval of data use. Efficiency is another consideration, with some states choosing to use a centralized data warehouse to better facilitate access to the data for specific purposes. More than twice the number of states use a centralized data warehouse for their SLDSs than use a federated structure.¹²

“When all these data are in one place, and agencies are just sharing them once, the agencies are not having to share data multiple times for these one-off projects that the legislature asks them to do. And what that means is less personal information being shared around the state to do this project or that project.”

– *Melissa Beard, data governance coordinator, Washington Education Research and Data Center*

In **Texas**, the Texas Education Agency, Texas Higher Education Coordinating Board (THECB), and Texas Workforce Commission all contribute data to a centralized data warehouse maintained by the THECB. The THECB de-identifies the data, assigning a unique identifier so that students can be linked, and therefore tracked, across systems. A data advisory board was established and consists of members of all three participating agencies. Linked data are sent to three Education Research Centers (ERCs), where researchers who have been approved by the advisory board can physically go to access the data. Commissioners from the participating agencies can request to access data at the ERCs for agency-supported research projects without approval by the advisory board but, as with external researchers, they must provide the resources (staff and money) to do so. The THECB has developed a comprehensive website with links to reports and data dashboards aimed at policymakers, schools and colleges, and students and their families. Data-sharing agreements across the three agencies allow for certain longitudinal data analyses to be conducted by agency staff and included in web applications and reports.¹³

The national experts and state education officials we spoke with suggested there is no one best model for an SLDS; rather, they said the best fit for a given state is what matches its history, culture, and capacity. However, they identified several key factors that are crucial in contributing to an effective and useful statewide data system:

- participation, at a minimum, of K-12 schools, public postsecondary education institutions/systems, and the state workforce agency; although many states include additional agencies—such as early learning, criminal justice, public health, and social services—to expand the issues that can be studied and addressed with the data;
- transparency about data security, access, and use; and
- legislation to formalize the structure and ensure compliance and continuity across changes in leadership.

Uses and Benefits of an SLDS

Regardless of the structure they choose for their data systems, states are finding SLDSs to be valuable tools in helping state policymakers, education institutions and systems, and students and families.¹⁴ The following sections outline examples of possible benefits for each of these stakeholders.

State policymakers. Having access to good statewide data enables policymakers to plan strategically and to address state needs. According to Christina Whitfield, senior vice president at the State Higher Education Executive Officers (SHEEO) organization:

At the state policy level, [policymakers need to] articulate a public agenda for higher education in the state and then put in place ways to measure progress toward [their] goals. Thinking about student success from a global or statewide perspective certainly is one of the key advantages of these [data] systems.

We found several examples of how states have used the data to address key policy issues:

- In Tennessee, the education data system helps the state monitor the effects of the Tennessee Promise policy. For example, the system helps the state track the progress of individuals through the Promise's intake process, to identify points where people drop out and to inform efforts to improve the process.
- In Texas, the statewide data system has been critical for strategic planning for the state's 60x30TX initiative. For example, the system helps to project the number of credentials the state's institutions need to produce in order to meet state goals, making it possible to break down the projections by students' race/ethnicity, gender, and economic disadvantage.

Education institutions and systems. Interviewees said that having access to cross-sector data has helped education institutions and systems to understand how well their policies and programs are working for students. The institutions have used this information to support improvements:

- In Minnesota, an interactive website allows school districts to observe, by various demographic characteristics, how well their graduates fare in postsecondary education, including in college enrollment, participation in remediation, and completion outcomes.¹⁵
- Community college systems in several states, including Tennessee and Virginia, have used analyses of linked data as a basis for making changes to their placement policies and developmental education offerings, with early evidence suggesting large declines in remedial placement and increases in successful completion of college-level gateway courses.^{16, 17, 18}

“The states with the strong [data] systems can put in place creative policies and look at them fairly quickly and tweak them so they’re not wasting a lot of time and money. Being able to have a good, robust data system and make adjustments on the fly is a huge benefit.”

– Patrick Lane, project manager for the Western Interstate Commission for Higher Education’s Multistate Longitudinal Data Exchange

Agencies in **Virginia** share data through a federated model, where the participating entities (K-12, higher education, workforce agencies, and social services) have a data-sharing agreement, but each agency controls its own data to its own standards, with no central data warehouse. A federal SLDS grant was used to start the data-sharing effort, and legislation codified the role of the State Council of Higher Education for Virginia (SCHEV) as administrator of the project. A support vendor performs a data match twice a year, assigning a new common identifier that is then cross-walked across agencies. Every participating agency has a member on the data governance council and has a say in what information is shared. Agreement on data sharing must be unanimous. External researchers can gain access to shared data, which is de-identified prior to release, but need approval from each participating agency whose data are used.

Students and families. In addition, states are using their SLDSs to provide information to students, families, and the general public to support informed college choices and a better understanding of the value of the states’ investment in education:

- In Texas, the SLDS feeds information into the Texas Consumer Resource for Education and Workforce Statistics, or CREWS, an interactive dashboard that displays information about aspects of two- and four-year postsecondary programs, including students’ time to degree and, for graduates, average wages and student loan debt.¹⁹
- In the state of Washington, the Education Research and Data Center conducted usability studies to determine how best to present data for parents and students to access. Informed by those studies, ERDC translates research findings based on its SLDS into useful information for the public.

Challenges in Creating an SLDS

The experts we spoke with described the challenges of developing SLDSs in other states in ways that mirror those we heard about in California from both state policy staff and local education leaders. They all agreed that California's biggest challenge is the lack of any organization responsible for cross-system planning. This leadership deficit was raised by state policy staff in California when describing the state's past efforts to develop a cross-sector data system. It also was cited by local educators who said the lack of any regional cross-sector organization was impeding their efforts to share data. Other challenges faced by states in developing an SLDS include:

- the need for more understanding about the value and purpose of cross-system student data;
- the need to build relationships and trust across education systems;
- limited resources and capacity for building and using better data systems;
- technical hurdles associated with combining information from different data systems; and
- the difficulty of sustaining efforts when there is turnover among top leaders and changes in resources and priorities.

New Jersey's approach to data sharing is more of a hybrid model, with data (from the State Education Department, Office of the Secretary of Higher Education, and Office of Labor and Workforce Development) held centrally at Rutgers University's Heldrich Center. A memorandum of understanding among the three participating entities allows the center to collect and link data, and maintain a centralized warehouse. But each party with data being used must approve any request by center staff to use data for research purposes. Center staff are the only ones with access to the linked data. Startup funds for the shared system came from a federal SLDS grant and a federal Workforce Data Quality Initiative grant; funding for ongoing maintenance is currently included in the higher education and labor budgets.

The technical barriers to creating an SLDS designed for research to support state policy and inform programs and practices at schools and colleges are minor, all interviewees (in California and other states) agreed, compared to the political and governance challenges. For example, K-12 and postsecondary data systems across the country use different student identifiers, but experts said solutions have been found to allow matching across these systems.

There are, however, significant technical challenges to developing the kind of real-time data system that some stakeholders in California told us they would like to have—one that allows for matching data across education institutions for use by programs that provide students with direct services, such as counseling and advising.²⁰ These kinds of uses require a powerful system that accesses data in real time, rather than a more limited system that collects and provides access to new data on a set schedule, such as annually or after each term. The experts we interviewed said that developing such a system would pose significant financial and technical challenges. In California, only the CSU system uses a uniform student information system platform across its campuses. There are a variety of platforms in use across UC campuses, community colleges, and K-12 districts, making it technically difficult and likely very costly in the near term to create "live" connections across all institutions that could be accessed by schools and colleges to provide counseling and other services to students. Without such live connections, schools and colleges would have to upload data very frequently (perhaps weekly or more often) to approximate a real-time data system, which is likely beyond the capacity and resources of most institutions.

How Would the Various Options Work in Considering an SLDS in California?

The California Context

The national and state experts we spoke with emphasized the importance of taking into account a state's history and culture when considering the development of an SLDS. Our research across this series points to several issues of importance for California:

- **Interest in cross-sector information among educators.** There is great interest among educators in California in accessing cross-sector data. This is evidenced by the voluntary participation of some school districts and postsecondary institutions in Cal-PASS Plus, the California College Guidance Initiative, and local/regional data-sharing efforts, despite the costs and challenges of doing so, as we described in the second and third briefs in this series.
- **Underutilized analytical capacity at research centers and universities.** California has a number of well-respected applied research centers that focus on education, such as Policy Analysis for California Education, the Higher Education Center of the Public Policy Institute of California, the Research and Planning Group of the California Community Colleges, and others. With better access to student-level data across education systems, this analytical capacity could be better used to support education policy and practice in the state.
- **No cross-sector entity.** The lack of a higher education coordinating board or other agency with a cross-sector mission was cited by interviewees as California's biggest challenge in creating an SLDS, since such entities often serve as the agency that develops and maintains SLDSs in other states. There have been several legislative efforts in recent years to create such an entity in California, but the bills failed to make it through the legislative process or were vetoed by the governor.²¹
- **History of resistance and lack of incentives.** Legislative efforts to create a mechanism for connecting data across the education sectors have generally met resistance by the state's four education systems. This is particularly the case for those efforts that would have created an agency with a role in coordinating the postsecondary systems. However, recent efforts by the systems to make better use of their own data to understand student progress have led to a greater recognition of the benefits of cross-sector data. This is evidenced by the voluntary creation of data-sharing agreements among the four system offices.²² Legislative efforts have also faced resistance from some in the policy community over concerns about feasibility and cost.

A Set of Policy Criteria to Assist in Considering an SLDS

The findings across the reports in this series suggest that California policymakers and educators should consider a set of values or criteria to facilitate decision-making concerning the feasibility and options for developing an SLDS in California. Primary among those values should be a focus on the public good—that is, on the purposes and benefits of sharing and using cross-sector data to understand and improve student outcomes through better-informed education policy and practice. The interests of education institutions and systems, while important, should not take precedence over California's need for greater transparency and for increased access to information to improve educational outcomes. Other important criteria to consider include data security, political and technical feasibility, cost, and sustainability (see Table 1).

Table 1

A set of values should guide any consideration of an SLDS in California.

Value/Criterion	Definition
Public Good	The mission is focused on using cross-sector data to produce information of value to policymakers, educators, students/families, and the public in a transparent way, for the purpose of improving student learning and other outcomes, including equity goals.
Data Security	Data are kept secure to ensure the privacy of individual students and to maintain compliance with state and federal laws.
Data Quality	Information in the data system is complete and accurate and relevant for addressing issues related to student success.
Cost and Time	The data system can be developed, maintained, and used in a reasonable timeframe and at a reasonable cost to provide a good return on taxpayers' investment.
Technical Feasibility	The data system relies on well-established technology and allows for incorporating technical improvements over time.
Political Feasibility	The data system and resulting information/analyses can achieve and maintain the support and trust of various stakeholders.
Sustainability	The data system can be maintained through changes in leadership at the state, system, and institutional levels.

Assessing the Options

Centralized or federated? Most states we examined use a centralized data warehouse model for their SLDSs. A data warehouse is more efficient to access and use once it is established. With a federated system, participating agencies must run data matching procedures and extract and load datasets into a central location for each individual use of cross-sector data. This requires more investment of time and resources on the part of those agencies, compared to the regular data submission schedule of a centralized data warehouse. These individualized processes also make the data more difficult to access, which has implications regarding public access and use (that is, the public good) and sustainability (due to changes in staffing and resources at the participating agencies). In addition, a centralized model could incorporate the ongoing work of the California Workforce Development Board to create a longitudinal data system for workforce education and training programs in order to provide a more complete picture of educational progress and outcomes in California.²³

Data governance. Interviewees at the local, state, and national levels said that determining where the data are housed and managed is one of the biggest challenges to creating an SLDS in California. We found four basic options for housing and managing a statewide cross-sector data system in California, whether that system is centralized or federated:

1. A new education coordinating body (either for postsecondary or for K-20 education) that could facilitate cross-sector planning and provide policy advice to the governor and legislature.
2. A new state data agency with no role in coordination or policy, or a new office for education data within an existing state agency.
3. A joint powers authority (JPA) between the systemwide offices of the four public education systems and the state.
4. A third party entity, such as a research center or other nonprofit organization.



Each option has benefits and downsides, and in our Assessment of Data Governance Options (see below), we compare the four options in relation to the set of values and criteria in Table 1. As the assessment suggests, a state coordinating body or a state data agency/office would appear to most clearly address the primary value of ensuring that the data system is developed and used to serve the public good. These options would also likely be the most trusted by the policy community and other stakeholders outside the education systems and could be the most sustainable once implemented (and funded) through legislation. With no role in coordination or planning, a state data agency/office would likely gain more support from the education systems.

A JPA would likely engender the least resistance from the education systems and could be more quickly implemented without going through the legislative process. Under a JPA, however, the education systems would maintain control over all decisions about the use of the data and any analyses, which could affect access to and use of the data system for research. Changes in system leadership could affect the priority assigned and the resources allocated to this cross-sector role, given the system offices' primary responsibility to serve the needs of their individual systems and institutions. In fact, national experts emphasized the importance of formalizing a data system through legislation, regardless of its governance and structure, to ensure compliance and sustainability.

Interviewees in states that use a university, a research center, or other third party to manage their SLDSs noted an advantage in the perceived independence of such an organization. However, it is unclear which state agency in California would manage the contract for such an organization, and the need for contract renewals over time could reignite debates over which entity is selected or other details about the design or use of the data system, raising questions about sustainability.²⁴ Some organizations with relevant expertise and experience to serve in this role have close relationships to a particular education system, which could make choosing an entity difficult due to real or perceived conflicts of interest.

Assessment of Data Governance Options

Value/Criterion	Considerations and Comparisons
Public Good	<ul style="list-style-type: none"> • A statewide coordinating body has a clear cross-sector mission, would be oriented toward transparent use of the data as part of that mission, and has potential capacity for translating findings into policy action through its role in advising policymakers. • A state data agency also has a clear cross-sector focus and an orientation toward use of the data as its primary mission. Without a role in planning or coordination, it has less ability to translate findings into policy action, but it could facilitate access to the data by applied researchers who could do analyses and make policy recommendations. • A JPA is less likely to adhere to a cross-sector mission and would likely be less oriented toward facilitating access to and use of the data for policy research due to conflicting priorities. • A third party could potentially have a clear cross-sector mission and could be charged with facilitating access to the data for policy-oriented research.
Data Security	<ul style="list-style-type: none"> • Keeping individuals' information secure in compliance with state and federal laws could be equally valued and implemented by each potential data governance option.



Value/Criterion	Considerations and Comparisons
Data Quality	<ul style="list-style-type: none"> Both a statewide coordinating body and a state data agency could have good leverage with participating agencies to ensure data quality. In a JPA, each system would understand quality issues with its own data. But the systems may have less leverage with each other than would a state agency to ensure data quality. A third party would have little leverage with the systems to ensure quality of data.
Cost and Time	<ul style="list-style-type: none"> Both a statewide coordinating body and a state data agency would take time to establish, although either entity could leverage existing state infrastructure (e.g., office space) to create some efficiencies. With either option, all staff and other resources would be dedicated to cross-sector work. A statewide coordinating body with a mission that includes planning and advising would cost more to establish and maintain than a state entity with a targeted data mission, especially if the data-focused entity were created as an office within an existing state agency. A JPA could be established fairly quickly and could leverage some infrastructure at the existing system offices; however, the existing staff at system offices have limited capacity to take on additional tasks, and any new staff and other resources may not be assigned solely to cross-sector data work. A third party entity may be able to leverage some of the state’s prior investments (such as for Cal-Pass Plus), but the state would have less direct control over its costs.
Technical Feasibility	<ul style="list-style-type: none"> Each governance arrangement could navigate the technical challenges involved in developing an SLDS. A statewide coordinating body or state data agency would allow the state to exercise the most control over technology investments over time. A JPA would be subject to the technology investment choices of the individual system offices; a similar situation would exist with a third party entity, though it might be nimbler in adapting to changing technology.
Political Feasibility	<ul style="list-style-type: none"> Both a statewide coordinating body and state data agency would likely face resistance from the education systems, but the systems would be significantly more resistant to a coordinating body with responsibilities other than data. Policymakers and external stakeholders would likely have the most trust in a non-political state data agency. A JPA could be formed without going through a legislative process; however, policymakers and external stakeholders are likely to have less trust in this arrangement. It also may preclude using a centralized data warehouse model if the education systems cannot agree on which of them would manage it. The political feasibility of a third party entity seems uncertain, because it is not clear which statewide entity would manage a contract. In addition, some potential external entities have close relationships with particular education systems, creating the potential for real or perceived conflicts of interest.
Sustainability	<ul style="list-style-type: none"> Once created (and funded), both the statewide coordinating body and the state data agency, with their cross-sector mission, would persist unless specifically eliminated by state policy. A JPA, because it involves the continued commitment of all signatories, is more subject to changes in leadership at the system level. New leaders could change the priority assigned and resources allocated to the cross-sector role. The difficulty in determining which state entity manages the contract for a third party entity, as well as the need for periodic contract renewals, could open up debates about the selection of the entity, design of the system, and other issues that could interfere with the work.



Costs. Many California stakeholders interviewed for our earlier briefs suggested that costs could be a significant barrier to the development of an SLDS in California. But our interviews of national experts and officials in other states suggest that the annual operating expenses of a data system similar to those that many other states have developed may not be as costly as some assume. The state representatives we spoke with estimated their annual costs to maintain an SLDS (including analysts, researchers, committees to approve research use, and other operating costs) at \$600,000 to \$1 million. Since California is the most populous state and would need additional researchers and technicians, those we spoke with estimated that the annual cost to California would likely be around \$2 million.

While a complete analysis of costs is outside the scope of our research, this estimate appears reasonable in the context of the operating costs of the former California Postsecondary Education Commission (CPEC), which collected and compiled data from the various higher education systems. The annual cost in current dollars of a new state agency with staffing and operating costs similar to what CPEC had at the time it was de-funded in 2011 would be approximately \$2.3 million.²⁵ An office within an existing state agency could achieve some savings in the form of lower administrative costs, leaving room for investment in additional technical or analytical staff.²⁶ An annual budget in the range of \$2 million to \$3 million would represent a small fraction of the state's annual investment of general funds in higher education and in K-12 schools (\$17.7 billion and \$54.1 billion, respectively, in 2017-18).²⁷

In addition to operating costs, there would be startup costs that are more difficult to estimate. Interviewees indicated that the initial technology costs to establish a data warehouse would be reasonable, noting that the price of servers, memory, and cloud computing services have declined in recent years. Staff time would make up the bulk of initial costs for the entity assigned with developing the SLDS, and at the participating agencies (CCC, CSU, UC, and the Employment Development Department). Additional work would include engagement in planning and decision-making around processes for data collection, data security issues, access to and use of the data, and other issues. The federal SLDS grant program might offer some guidance as to these costs, as many states have used these grants to establish their SLDSs; most grants have been less than \$10 million, though some have been in the range of \$10 million to \$20 million.²⁸ However, many states used these grants to build longitudinal data systems within a single sector (like California did in creating CALPADS) and to cover operating costs for their SLDSs during the early years, before building those ongoing costs into state budgets. California already has strong student-level data systems within each education sector and could make plans for covering annual operating costs from the start, keeping the state's initial costs lower than some federal SLDS grants might suggest.

There would likely be some cost savings to offset some of the state's investment in developing and maintaining an SLDS. For example, each higher education system in California, along with many school districts and individual college campuses, currently pay for data matches to the National Student Clearinghouse (NSC) as a means of tracking students' postsecondary enrollment and completion. These expenditures, generally made from state funding allocations, could potentially be consolidated into a single contract by the entity managing an SLDS, likely at a reduced cost per student matched. In addition, while matches to the NSC would remain useful in order to observe postsecondary enrollments and completions of California students in private and out-of-state institutions, an SLDS that included the public postsecondary systems would capture most of the activity that our institutions currently have to contract with the NSC to observe.²⁹

Recommendations

California Needs a Statewide Data System

Our research findings across this series point to the need for, and potential value of, a statewide longitudinal education data system in California. The current student data systems at the four public education systems, while containing good information about portions of the student journey, are disconnected, inaccessible, and underused, leaving important questions unanswered. Regional data sharing, while adding value and providing a platform for collaboration across institutions, is not an efficient or effective substitute for a statewide data system. California lags behind many states that are using SLDSs to provide critical information to policymakers and educators for improving policy and practice and to provide information to help students and their families make educational choices.

In Managing an SLDS, a State Data Agency/Office Aligns Best with Public Needs

In considering the various governance options for an SLDS, the creation of a new state data agency or office within an existing state agency would align best with the set of values set forth for a state data system in California. Such an agency/office could have the mission of developing and managing an SLDS, creating standard reports and data dashboards for various audiences (policymakers, schools and colleges, students/families), and managing access to the data by external researchers. While the data agency/office itself would not have a role in planning or coordination, these external researchers could analyze the data and make recommendations for improvements to education policy and practice. Creating an office within an existing agency—such as the California State Library, the Senate Office of Research, the Department of Finance, or the Governor’s Office of Planning and Research—would have some administrative cost savings. Creating a separate state agency might have other advantages, in terms of keeping its mission (and staffing) clearly insulated from any existing agency’s mission and politics.

“For California, it’s critical to weigh the cost of not using data, with equity considerations not just [among] students, but [among] regions of your state, where some educators and families have this information and some educators and families don’t. We’ve heard leadership in Sacramento say, ‘This is best left to locals, let locals decide, they don’t need Sacramento in their business.’ That makes sense to a point; you can champion local decision-making, but you can actually enable it by ensuring equitable access to the information [people] need to make those decisions.”

– Paige Kowalski, executive vice president, Data Quality Campaign

In Terms of Structure, a Centralized Data Warehouse Aligns Best with Public Needs

Regardless of whether data governance is through a new agency or an office within an existing agency, it appears that developing a centralized data warehouse would be a better option than managing a federated process for data sharing. A data warehouse model would better facilitate access to and use of the information to inform education policy and practice. It would also be less burdensome for the education systems, which would have to upload data on a set schedule (such as annually, or after each term), rather than with each individual use of cross-sector data.



The Creation of an Effective SLDS Does Not Require a New Coordinating Body

There have been many calls for California to create a new coordinating body to facilitate planning across the CCC, CSU, and UC systems.³⁰ The benefits of greater coordination go beyond the integration of data across sectors to include stronger policy leadership, more effective planning, greater accountability for educational outcomes statewide, and better alignment of funding, policies, and programs across education sectors.³¹ While California would certainly benefit from a higher education or K-20 coordinating body, the political hurdles to creating such an agency have been daunting. The need for better cross-sector data to support improvements in education policy and practice is urgent and can be met without a coordinating body. While some states house their SLDSs in a higher education coordinating body, others assign the management of their SLDSs to a state data agency/office. For example, Washington has a coordinating body—the Washington State Higher Education Coordinating Board—but houses its SLDS in the Education Research and Data Center within the governor’s budget agency. A new gubernatorial administration in California may present opportunities to create a coordinating body and consider whether to house the SLDS in that agency. However, developing a data system does not have to be contingent on resolving the issue of whether to create a coordinating body.

Possible Steps to Implementation

Legislation would be required to create a new state data agency/office that would be charged with developing an SLDS. This legislation, in defining the agency’s structure and authority, could:

- describe the purpose of the agency/office as being to collect, compile, and disseminate data to state parties, such as the legislature and executive agencies, and to educational institutions, external researchers, and the general public;
- specify that the agency/office would develop an SLDS and facilitate access to the data by researchers and practitioners to inform education policy and practice; and
- define a governance council, which could include agency/office staff, a representative from each of the participating agencies (the four public education systems and the Employment Development Department), and several other relevant stakeholders (such as the Legislative Analyst’s Office, Department of Finance, Senate Office of Research, and/or other relevant state agencies).³²

Once created, the new state agency/office and its governance council would need to enact regulations to more clearly define the processes for collecting data, such as specifying the kinds of data that participating agencies would be required to submit and on what schedule. Regulations would also clarify the processes for use of information in the SLDS and could include:

- a requirement that any cross-sector question asked of the new agency/office by the legislature or a state executive agency must be answered (given data are available);
- a description of the kinds of data reports and dashboards the agency should create, including information aimed at K-12 schools (such as information on college applications, college enrollment, and course placement of graduates by high school), postsecondary institutions (such as transfer outcomes by community college and employment outcomes by postsecondary institution, degree, and major), policymakers (such as common accountability metrics), and students (such as information on time to degree and student loan debt);
- a requirement that reports and dashboards disaggregate data for major student populations (such as by race/ethnicity and income) and by region of the state;
- a process whereby education institutions and their systemwide offices could extract de-identified student records to do their own analyses;



- a process for external researchers to gain access to de-identified student records for research, subject to review and approval of their proposal by a committee and limited to questions addressing issues that serve the public interest and require data from multiple education systems; and
- a requirement that the agency catalog reports of studies that are conducted using the data and otherwise facilitate dissemination of the knowledge gained through this public resource.³³

Limitations of this Approach

An SLDS structured as we describe here would provide valuable information for research to inform improvements in education policy and practice and to increase transparency and understanding around education pipeline issues. Schools and colleges could gain valuable information about their students and programs through this data system. For example, the state data agency/office could develop dashboards that show metrics related to student progress and outcomes on an annual basis, allowing schools and colleges to monitor changes in these metrics from one year to the next. And institutions could extract data sets to conduct their own analyses to address local priorities. However, this kind of data system would not provide schools and colleges with real-time information to serve the immediate needs of current students. As described earlier, a real-time data system on a statewide scale is more technically challenging and costly.

In addition, we have focused only on the public education sectors in this report and in our recommendations. While it would be ideal to include student-level data from private institutions in an SLDS, an initial focus on the public education sectors would be reasonable and would cover a significant majority of student enrollment in California. Adding data from private institutions, and from state agencies involved in health, social services, or other relevant issues, could be considered over time.

Final Thoughts

In recent years, California has enacted a raft of new policies and practices in both K-12 and higher education, including the Local Control Funding Formula and Local Control and Accountability Plan in the K-12 system, the Student Success and Support Program and the Associate Degree for Transfer in the community colleges, the CSU's Graduation Initiative 2025, and the new UC Transfer Pathway plan. Policymakers have created a number of mechanisms to encourage cross-sector collaboration in the pursuit of better educational outcomes, including the Career Technical Education Incentive Grant and the Awards for Innovation in Higher Education. It is imperative for the state to be able to understand the success of its various reforms and initiatives. It is equally important that school districts, community colleges, and universities are able to assess and improve their programs, which requires access to information about their students' prior educational experiences and their students' success in further education and the workforce. Both state policymakers and local educators need access to adequate information to carry out their roles to improve student learning, progression, and success, and it is time for California to help by developing a vital tool to support their work.

Acknowledgments

We thank the national experts and state officials who participated in the interviews conducted for this report, including John Armstrong, Melissa Beard, Elizabeth Dabney, Julie Eklund, Meredith Fergus, David Grenier, Stephanie Huie, Paige Kowalski, Kathy Krepcio, Patrick Lane, Tod Massa, Mamie Voight, Margaret Votta, and Christina Whitfield. Their generosity in sharing their time and expertise was essential to this research. Amy Supinger assisted us in understanding CPEC's operating costs and the implications for a new state data agency/office, and Glen Wilkins provided research assistance related to data systems in other states. We appreciate the participation of key colleagues in discussions to help clarify our thinking around key findings and recommendations, including Jorge Aguilar, Kathy Booth, Pamela Burdman, Hans Johnson, Thad Nodine, Patrick Perry, Robert Shireman, and Amy Supinger. We are grateful for Thad Nodine's guidance in structuring and framing the content of the report, and for the additional review and comments provided by Lande AJose, Kathy Booth, Pamela Burdman, Sebastian Castrechini, Jacob Jackson, Christian Osmeña, Patrick Perry, and Andrea Venezia, which helped us improve an earlier draft. We appreciate the editorial assistance of Gretchen Kell and Sasha Horwitz and the graphic design work of the Foundation for California Community Colleges. The authors take responsibility for all content.



Endnotes

- 1 Moore, C., Grubb, B., & Esch, C. (2016). *Gaps in perspective: Who should be responsible for tracking student progress across education institutions?* Sacramento, CA: Education Insights Center, California State University, Sacramento.
- 2 Moore, C., Bracco, K. R., & Nodine, T. (2017). *California's maze of student information: Education data systems leave critical questions unanswered.* Sacramento, CA: Education Insights Center, California State University, Sacramento.
- 3 Moore, C. & Bracco, K. R. (2018). *Scaling goodwill: The challenges of implementing robust education data sharing through regional partnerships.* Sacramento, CA: Education Insights Center, California State University, Sacramento.
- 4 This definition would not include all data systems developed with funding through the federal SLDS grant program operated by the National Center for Education Statistics. For example, the CDE received an SLDS grant in 2006 to create the California Longitudinal Pupil Achievement Data System (CALPADS), which tracks students through the state's K-12 system. Since it does not follow students into postsecondary education or the workforce, it would not meet our definition.
- 5 Perez, Z. (2016). *50-State Comparison: Statewide Longitudinal Data Systems.* Denver, CO: Education Commission of the States. Retrieved from <https://www.ecs.org/state-longitudinal-data-systems>. California is not included among the 37 states with an SLDS. While California's education systems have occasionally matched student records for a particular purpose, there are no ongoing "formal connections" between the systems, as required to meet the Education Commission of the States' definition of an SLDS.
- 6 Armstrong, J. & Whitfield, C. (2016). *The state of state postsecondary data systems.* Boulder, CO: State Higher Education Executive Officers Association. Retrieved from http://www.shereo.org/sites/default/files/publications/SHEEO_StrongFoundations2016_FINAL.pdf.
- 7 We interviewed national experts at the State Higher Education Executive Officers Association, the Institute for Higher Education Policy, the Western Interstate Commission for Higher Education, and the Data Quality Campaign. We also interviewed state officials involved in SLDSs in Minnesota, New Jersey, Rhode Island, Texas, Virginia, and Washington. States were selected based on national experts' recommendations of states that represent a range of approaches to the structure and management of their SLDSs.
- 8 Armstrong and Whitfield, 2016. For more information on the Statewide Longitudinal Data Systems Grant Program, including publications on best practices in the development and use of SLDSs, see <https://nces.ed.gov/programs/slds>.
- 9 Information on the Data Quality Campaign can be found at <https://dataqualitycampaign.org>.
- 10 For more discussion about state considerations in developing and using an SLDS, see: Perez, Z. (2017). *Examining SLDS development and utility.* Denver, CO: Education Commission of the States. Retrieved from https://www.ecs.org/wp-content/uploads/Examining_SLDS_Development_and_Utility.pdf; and Armstrong, J. & Zaback, K. (2016). *Assessing and improving state postsecondary data systems.* Washington, D.C.: Institute for Higher Education Policy. Retrieved from http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/state_postsecondary_data_systems.pdf.
- 11 For a more complete description of centralized and federated models for an SLDS, including the pros and cons of each approach, see Institute of Education Sciences (2012, October). *Centralized vs. federated: State approaches to P-20W data systems. SLDS Issue Brief.* Washington, D.C.: National Center for Education Statistics. Retrieved from https://nces.ed.gov/programs/slds/pdf/federated_centralized_print.pdf.
- 12 Perez, 2016. See a list of states using each approach at <http://ecs.force.com/mbdata/MBquestnb2?Rep=SLDS1604>.
- 13 See www.txhighereddata.org. For a description of the various education data resources in Texas, see Texas Higher Education Coordinating Board (2018). *Overview: Higher education data resources.* Austin, TX: Author. Retrieved from <http://www.theccb.state.tx.us/reports/PDF/10443.PDF?CFID=76207717&CFTOKEN=45100928>.
- 14 For more information on states' uses of SLDSs, see Armstrong, J. & Whitfield, C. (2016).
- 15 The website for the Minnesota Statewide Longitudinal Education Data System (SLEDS) can be found at <http://sleds.mn.gov>.
- 16 Edgecombe, N. (2016). *The redesign of developmental education in Virginia.* *New Directions for Community Colleges*, 2016(176), 35-43.
- 17 Denley, T. (2016). *Co-requisite remediation full implementation 2015-16. Tennessee Board of Regents Technical Brief No. 3.* Retrieved from <https://www.tbr.edu/sites/tbr.edu/files/media/2016/12/TBR%20CoRequisite%20Study%20-%20Full%20Implementation%202015-2016.pdf>.
- 18 Scott-Clayton, J. (2018). *Evidence-based reforms in college remediation are gaining steam – and so far living up to the hype.* Washington, D.C.: Brookings Institution. Retrieved from <https://www.brookings.edu/research/evidence-based-reforms-in-college-remediation-are-gaining-steam-and-so-far-living-up-to-the-hype>.
- 19 See <http://www.theccb.state.tx.us/apps/txcrews>.

- 20 As one example of this kind of use, community college counselors meeting with high school seniors could access the students' high school records to see what classes they have taken, their grades, and other information for use in course placement and in helping the students design an education plan.
- 21 A current bill, AB 1936 (Low and Eggman, 2017), would create an Office of Higher Education Performance and Accountability, with part of its role being to develop and maintain longitudinal student data. The legislation is essentially a reintroduction of several previous versions of the bill that have been put forward nearly annually since 2012.
- 22 These data-sharing agreements have not been widely used, and the results of any analyses are not necessarily shared with the policy or education communities. The agreements also do not address the issue of access to cross-sector data by researchers outside the education systems.
- 23 As part of the Workforce Innovation and Opportunity Act (WIOA), the federal government required states to develop accountability dashboards covering adult education, career and technical education in community colleges, and other workforce training programs offered through Workforce Investment Boards. In response, California Assembly Bill 2148 (Mullin, 2014) called for the development of an annual workforce metrics dashboard and authorized the Employment Development Department to collect participant data for workforce training programs and provide the data to California Workforce Development Board for use in the dashboard. The text of AB 2148 is available at http://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201320140AB2148.
- 24 The state's experience with the Cal-PASS Plus system, contracted through the CCC Chancellor's Office, offers lessons in this regard, as the contract for managing that data system has changed hands over time, affecting the continuity and character of the work being done with those data, as well as the choices of institutions about participating in the effort. Some K-12 and university interviewees noted that Cal-PASS Plus is often viewed in their sectors as a resource run by and for the community college system, which may contribute to the limited participation by K-12 districts and UC/CSU campuses.
- 25 This figure was calculated by applying the current monthly salary (using the midpoint of the range for each classification, plus 50.4% for benefits) to positions representing the same number and type of staff that CPEC had in 2011, with the addition of a high-level data manager. We also added \$25,000 per employee to cover initial costs for workspace, computer, etc. We included an additional \$900,000 for non-staff-related operating costs, representing an annual 3 percent increase over CPEC's operating cost in 2008-09 of \$650,000 (after rounding).
- 26 We estimated that removing the executive office staff and reducing the administrative staffing from three positions to two would reduce the annual cost from \$2.3 million to \$1.8 million.
- 27 Figures drawn from the 2017-18 state budget, enacted budget summary, chapters for K-12 Education and Higher Education, retrieved from <http://www.ebudget.ca.gov/budget/publication/#/e/2017-18/BudgetSummary>.
- 28 Information on SLDS grants received by states since 2006 can be found at <https://nces.ed.gov/programs/slids/stateinfo.asp>.
- 29 According to the National Center for Education Statistics, about one in 10 California residents enrolling as first-time postsecondary students go out of state (*Digest of Education Statistics 2016*, Table 309.10). About 18 percent of California's total postsecondary enrollment in fall 2015 was in private non-profit or for-profit institutions (Table 304.60), with the remaining 82 percent enrolled in the CCC, CSU, and UC systems.
- 30 Examples include: Jackson, J. & Johnson, H. (2018). *California's future: Higher education*. San Francisco, CA: Public Policy Institute of California; and California Competes (2015). *Mind the gap: Delivering on California's promise for higher education*. Berkeley, CA: Author. In its most recent report, California Competes called for the creation of a higher education coordinating entity, with collecting and reporting on cross-sector data as one of its roles. See California Competes (2018). *Out of the dark: Bringing California's education data into the 21st Century*. Berkeley, CA: Author.
- 31 Heiman, J. (2010). *The Master Plan at 50: Greater than the sum of its parts—Coordinating higher education in California*. Sacramento, CA: Legislative Analyst's Office.
- 32 The legislation would also need to change the California Education Code to eliminate CPEC, which is still assigned the role of collecting cross-sector data from the various higher education segments (CPEC was de-funded in 2011, but not eliminated).
- 33 For example, the Texas Education Research Center includes links to policy briefs and other publications based on researchers' use of its SLDS. See <https://texaserc.utexas.edu/about-us/publications>.

