# Developing a Comprehensive Data System to Further Continuous Improvement in California

Julia Koppich, Evan White, Simon Kim, Marcy Lauck, Noah Bookman, Andrea Venezia

Governor Gavin Newsom's budget proposal for 2019–2020 includes $10 million to develop a statewide longitudinal data system—including early education, K–12, and higher education institutions as well as health and human services agencies—to better track student outcomes and improve alignment of the education system to workforce needs. California's lack of a coherent education database serves as a substantial barrier to fulfilling the state's continuous improvement policy goal and ensuring all students have access to robust learning opportunities to enable them to be successful in school and beyond. This brief reports on a session of the PACE conference held on February 1, 2019, that brought together experts to discuss a set of essential questions California must consider as it develops a new coordinated data system.

California's new accountability and continuous improvement framework relies on district and school leaders using multiple measures of school performance to identify where change is needed, and to monitor carefully the development, testing, and evaluation of improvement strategies over time. This process of continuous improvement requires that local leaders have access to research-based evidence and strategies that they can implement in their schools and opportunities to learn from one another about what works, under which conditions, and for which students. PACE's series of Continuous Improvement Briefs aims to support education leaders at all levels in learning how to improve the performance of their schools and students.

PACE
*Policy Analysis for California Education*

May 2019

## Introduction

On January 10, 2019, Governor Gavin Newsom released his first state budget. The budget proposes a $10 million one-time allocation to develop a longitudinal data system "that would better track student outcomes and increase the alignment of our educational system to the state's workforce needs" (ebudget.ca.gov). The new data system would connect early education, K–12, and higher education institutions with relevant health and human services agencies, as well as with employers. Governor Newsom's May 2019 revised budget maintained $10 million for a statewide longitudinal "cradle-to-career" data system.

The new data system would connect multiple sectors of education and several agencies that, while they live outside of education, have a direct effect on students' ability to be successful in school. Such a coordinated data effort would, in the words of the Governor's proposal, "improve coordination across educational data systems and better track the impacts of state investments on achieving educational goals."

This brief reports on one session at the PACE annual conference, held February 1, 2019, in Sacramento. The session, "Promising Practices in the Development and Use of Cross-system and Cross-sector Data Systems," brought together experts with policy and practice experience to discuss a set of essential questions California must consider as it develops a new coordinated data system, including:

- What purpose(s) should the new data system serve?
- How should the new system be governed?
- Who should have access to it?

## Setting the Scene

Andrea Venezia of EdInsights, a policy and research center whose work focuses across the K–12, community college, and public university sectors, served as moderator and set the scene. She offered two main points, both of which would be echoed by the panelists who followed.

First, California cannot currently understand how successfully the state's public education systems are meeting students' needs or where gaps exist. The state, often a trendsetter, lags behind other states in terms of its ability to link data across systems. California is thus unable to answer key education policy questions that would be informed by collective data available in a more comprehensive and coordinated system.

An integrated data system that links student information from multiple sectors—early education, K–12, and higher education—and from multiple agencies—health, welfare, juvenile justice—could provide policymakers and practitioners a more comprehensive picture of students' paths from early education to and through the workforce and allow for the development of evidence-based interventions and system change initiatives to make it more likely that larger numbers of students would experience greater success.

Second, the work of constructing a multi-sector, multi-agency data system is not simply a technical exercise. It is "political and relational." California will need to determine what purposes this new integrated data system should serve and therefore which institutions and agencies ought to be part of it. The state must figure out how to bring these disparate entities together in this common effort and develop a structure suited to the identified purposes.

## Making California Data More Useful for Educational Improvement[1]

Evan White is the Executive Director of the California Policy Lab, a collaborative effort of researchers at UC Berkeley and UCLA that brings researchers together with state and local governments to generate data to bring to bear on issues ranging from homelessness to poverty, crime, and educational inequality.

White acknowledged that California has improved its data system, especially through the California Longitudinal Pupil Achievement Data System (CalPADS) that maintains individual student level data and enables tracking of each student's academic performance over time. But, he noted, "data availability and usefulness still fall short of other states." California is one of only six states in the nation that do not combine data across sectors. As a result, data are not well integrated. K–12 through CalPADS, early education, and higher education all maintain their own data systems, as do health and human services. These systems, White said, are only "lightly connected."

The result of, as White described it, California's "patchwork of data systems" is that the state is unable to answer its most important policy questions. A linked data system could ameliorate this problem by making available to a wide range of potential users—teachers and school and district leaders; the California Department of Education; higher education institutions; social service, employment, and law enforcement agencies; and researchers inside and outside government—the kinds

of data that could form the basis for evaluating multi-dimensional education policies and programs, identifying best practices, and providing an evidentiary foundation for more timely and on target student supports.

White identified three potential barriers to California's goal of linking data across sectors and agencies, while noting that all of these can be overcome. The barriers fall into three categories:

1. *Technical:* No common identifiers currently exist across all data systems. Privacy and security concerns are significant. Institutional capacity to link data in many cases is less than optimal.
2. *Bureaucratic:* Each agency owns its own data. Data governance is fragmented. No overarching governing body has the authority to grant or deny access to all of the data.
3. *Political:* Tensions around issues such as accountability, immigration enforcement concerns, and fears about what research may show can limit the development of a comprehensive system.

He suggested four near-term steps the state should take to reduce the likelihood that these barriers could stand in the way of a new data system:

1. Build on the state's existing infrastructure to develop data linkages. The state need not begin from scratch. The building blocks for a new integrated system lie within existing data systems.
2. Champion data integration through strong political and agency leadership. The state needs to send a clear message that articulates the value of an integrated data system and must work to engage stakeholders in building a collaborative system suited to the needs of multiple users.
3. Establish a governance structure that anticipates potential issues before they arise. This would include developing rules about access, use, privacy, and security to which all data users agree to adhere.
4. Determine where the system will "live." Deciding where data will be housed can impact accessibility and ease of use.

White concluded with a reminder that the challenges inherent in developing and implementing a comprehensive multi-sector, multi-agency data system have been met and successfully dealt with by other states. California would be well advised to learn from these experiences, taking from them what is useful and adapting lessons to the California context.

## Putting Evidence Into Action to Advance Equity in California: Building a Cross-sector Data Warehouse

The presentation by Simon Kim, Associate Vice President for Research and Sponsored Programs at California State University Long Beach (CSU Long Beach), detailed his experience developing a multi-sector data warehouse for the Long Beach community.

Long Beach's data warehouse is the result of collaboration across three institutions: the Long Beach Unified School District (LBUSD), Long Beach Community College (LBCC), and CSU Long Beach. The goal is to house and make the student level data available to the three institutions. This multi-sector data warehouse will make it possible to track student level data and allow the institutions collaboratively to identify risk and success factors that affect student access and achievement, thus enabling the Long Beach education community to make data-driven decisions to support students across the educational pipeline.

Kim described the genesis of the Long Beach warehouse. He was asked by the LBUSD school board to determine how graduates who attend CSU Long Beach were faring. Kim examined the available data and was able to deliver good news to the board. Trends were positive. Retention and graduation rates were increasing and equity gaps were closing. However, when asked to attribute these trends, he found he did not have sufficient data to answer the school board's question.

Three years ago, he secured a grant from the James Irvine Foundation to begin to create a single comprehensive data warehouse to include information about Long Beach students from the school district, LBCC, and CSU Long Beach. These collective data could help to answer the question posed by Long Beach school board.

Long Beach already had a history of cross-sector collaboration. The Long Beach College Promise, launched in 2008, is a commitment of the Long Beach Unified School District, LBCC, CSU Long Beach, and the City of Long Beach to guarantee an opportunity for a college education to every student who graduates from Long Beach Unified. As part of developing the data warehouse, the memorandum of understanding that governs the Long Beach College Promise was revised to add a commitment from the participants to share student level data. Kim described this arrangement as a "true partnership."

One of the key challenges to developing the coordinated data warehouse was that every participating institution

maintained a firewall for its data. It was not possible for the school district or community college simply to send data to CSU Long Beach where the database was being built. In effect, Kim needed to build a single data infrastructure integrating three different systems.

For example, each institution had a different way of coding data. Thus, matching students was a particular dilemma. As K–12 does not use social security numbers as identifiers, Kim used first and last names and dates of birth to match students, managing to achieve a 99 percent match.

Long Beach's data warehouse is now developed to the point where the institutions are beginning to merge data files. Once completed, the structure will represent the first time that Long Beach school district, community college, and CSU student level data will be in one place. Next, Kim hopes to develop a prediction model that can better explain and pinpoint factors that contribute to student success at each level of the Long Beach education pipeline.

Kim offered two cautions as California moves forward with a statewide longitudinal data system: patience and planning. First, it takes considerable time to align data elements. This process is complex and cannot be rushed. Second, the state needs to consider at the outset what the purposes of the database are and how the data will be used. Otherwise, Kim warned, there is a risk that considerable time and money will be expended developing a system no one will use.

## Building a Regional Cross-agency Data System for Education Improvement

Marcy Lauck, Director of Data Governance Strategies for the Santa Clara County Office of Education (SCCOE), described a set of data initiatives spearheaded by SCCOE. These initiatives offer a clear example of regional leadership in the data collaboration and coordination sphere.

The Silicon Valley Regional Data Trust (SVRDT) is led by the Santa Clara County Office of Education in partnership with the San Mateo and Santa Cruz county offices of education and health and human services agencies (juvenile probation, behavioral health, child welfare). The University of California at Santa Cruz serves as the initiative's research partner. The SVRDT, said Lauck,

> ...offers a secure data environment to enable public schools and health and human service agencies in San Mateo, Santa Clara, and Santa Cruz counties to legally and responsibly share data to coordinate case management, personalize and integrate services, inform public policy, and to partner with UCSC faculty to conduct research in partnership with public schools and county agencies.

DataZone, SVRDT's education warehouse for school districts, provides a range of services to 43 participating school districts including low-cost data warehousing, local data dashboards that display data that are updated nightly, and access to aggregated data from multiple sources (from education as well as health and human service agencies).

Professionals from a number of youth-involved agencies, including school services personnel, probation officers, and foster youth case workers, use data from DataZone's FosterVision application. FosterVIsion is a partnership between SCCOE, the Department of Family and Child Services, and Juvenile Probation to coordinate and improve services to foster and justice-involved youth. Currently 28 districts participate in FosterVision.

Finally, as part of its overall data coordination efforts, the SCCOE is engaged in a project to integrate early learning and K–12 data, which Lauck referred to as "crossing the great divide." The goal of this work is to create an infrastructure and governance arrangement for cross-agency data sharing between K–12 districts and early learning providers. The new data coordination effort uses the California Department of Education's Statewide Student Identifier (SSID)[2] as a common link across early childhood providers and school districts.

## Providing Educators a Vision of School Progress Through the CORE Data Collaborative

The final panelist was Noah Bookman, Executive Director of the CORE Data Collaborative. CORE was founded in 2010. A small group of California school districts combined their efforts around issues of academic standards. Today, eight districts—Fresno, Garden Grove, Long Beach, Los Angeles, Oakland, San Francisco, Sacramento, and Santa Ana—compose CORE.

The CORE Data Collaborative was established in 2015 when the CORE districts opened their data system to all interested school districts. Districts not among the CORE eight pay a fee on a sliding scale based on their number

of students. The Data Collaborative currently serves 80 districts and provides opportunities for educators across the state to share data, research, and innovations to help them gain a more precise picture of school progress in their district.

CORE was clear from its inception about the purposes of working with data. One of the Data Collaborative's foundational principles is that school quality must be about more than student achievement. It must encompass ways in which the school affects the whole child. Understanding the school's impact on its students requires careful attention to multiple measures, including students' social-emotional growth and indicators of school culture.

The CORE data system is built on the organization's beliefs about the purpose and use of data. The system includes state measures—student scores on standardized tests, English learner progress, chronic absenteeism, suspension rates, graduation rates—as well as CORE districts' locally driven measures—student academic growth, social-emotional learning, school culture and climate, high school readiness, and, still being developed, college and career readiness.

CORE Data Collaborative work focuses on using data to understand the growth of individual students in order to determine where the greatest impact on student learning lies in an effort to answer the question, "Is what we're doing working?" The Collaborative uses data to try to operationalize research results, turning these results into metrics that schools and districts can apply and use.

Bookman described a current piece of Data Collaborative research that aims to enable researchers to see data patterns that can help districts shape their approach to college and career readiness. CORE is using metrics that track students in Grades 3 through 12 to assess their progress towards graduation and career readiness. Data examined include assessments, behavior, attendance, course grades, and social-emotional learning survey responses. CORE is using these data to develop indicators that will predict early success in college and ultimately make it possible to backward map from these success indicators to enable K–12 districts to identify attributes likely to be related to student success.

According to early findings from this research, grade point averages and course-taking patterns are the strongest predictors of high school graduation and postsecondary outcomes. Persistence to a second year at four-year colleges is a strong indicator a student will graduate. Persistence to a second year at two-year colleges, however, is not a strong indicator of graduation.

Bookman offered a cautionary note for the state that mirrored Simon Kim's remarks. Specifically, design of the data system is critical. Too often system design is under-resourced. Developers spend too little time and money up front—neglecting key questions about who the system users will be and how they will use the data—and find themselves with a data system ill-suited to its purposes. Bookman reinforced what previous panelists had noted, namely, that California can learn from other states that have done data system development work and should take advantage of the opportunity to do so.

## Conclusion

This session offered presentations from several experts to help guide California as it moves towards an integrated multi-sector, multi-agency data system. Among the lessons to be taken from these presentations are that the state must carefully consider a range of issues, including the purpose of the system, the uses to which the data likely will be put, who will have access to the data, how the system will be governed, and where it will be housed. What also seems clear is that the state has much it can learn from regional and network data systems like the Silicon Valley Regional Data Trust and the CORE Data Collaborative.

At the end of the session during the question and answer period, the issue of the cost of building and maintaining a longitudinal data system in a state as vast as California came up. As one panelist said, "$10 million is a good start."

---

[1] This presentation was adapted from the *Getting Down to Facts* paper, "Making California Data More Useful for Education Improvement," by Meredith Phillips, Sara Reber, and Jesse Rothstein, found at www.gettingdowntofacts.com/publications/making-california-data-more-useful-educational-improvement.

[2] The State Student Identifier (SSID) is a unique, non-personally identifiable number linked to each K–12 student. The SSID makes it possible to maintain data on individual students, such as linking students to statewide assessment scores and tracking students in and out of schools and districts (cde.ca.gov).

## About the Authors

**Julia Koppich** is President of J. Koppich & Associates

**Evan White** is Executive Director at California Policy Lab

**Simon Kim** is Associate Vice President of Research and Sponsored Programs, California State University, Long Beach

**Marcy Lauck** is Director of Data Governance Strategies, Santa Clara County Office of Education

**Noah Bookman** is Executive Director of CORE Data Collaborative

**Andrea Venezia** is Executive Director of EdInsights Center, California State University, Sacramento

## Policy Analysis for California Education (PACE)

Policy Analysis for California Education (PACE) is an independent, non-partisan research center led by faculty directors at Stanford University, the University of Southern California, the University of California Davis, the University of California Los Angeles, and the University of California Berkeley. PACE seeks to define and sustain a long-term strategy for comprehensive policy reform and continuous improvement in performance at all levels of California's education system, from early childhood to postsecondary education and training. PACE bridges the gap between research and policy, working with scholars from California's leading universities and with state and local policymakers to increase the impact of academic research on educational policy in California.

Founded in 1983, PACE

- Publishes policy briefs, research reports, and working papers that address key policy issues in California's education system.
- Convenes seminars and briefings that make current research accessible to policy audiences throughout California.
- Provides expert testimony on educational issues to legislative committees and other policy audiences.
- Works with local school districts and professional associations on projects aimed at supporting policy innovation, data use, and rigorous evaluation.

## Related Publications


Heather Hough, Erika Byun, & Laura Steen Mulfinger. **Using Data for Improvement: Learning from the CORE Data Collaborative.** *Getting Down to Facts II*. 2018.

Meredith Phillips, Sarah Reber, & Jesse Rothstein. **Making California Data More Useful for Educational Improvement.** *Getting Down to Facts II*. 2018

Colleen Moore, Kathy Reeves Bracco, Thad Nodine, Camille Esch, & Brock Grubb. **California Education Policy, Student Data, and the Quest to Improve Student Progress.** Education Insights Center. 2018.


PACE
*Policy Analysis for California Education*


**edpolicyinca.org**

Stanford Graduate School of Education
520 Galvez Mall, CERAS 401
Stanford, CA 94305-3001
Phone: (650) 724-2832
Fax: (650) 723-9931