



# Out of the Dark:

## Bringing California's Education Data into the 21st Century

### Introduction

Higher education continues to be the linchpin for economic prosperity in California, serving as both an antidote and vaccine to income inequality. To significantly increase the number of Californians with a postsecondary credential or degree, our state needs a stronger grasp on not only students' experiences in the higher education system, but also their experiences with the systems that precede and follow. To improve higher education outcomes, California needs a clearer understanding of students' pathways from K-12 to postsecondary education to the workforce—an understanding that is based on data. With a new election cycle approaching and an increased public focus on data, the policy window is opening for California to catch up with the progress that most other states have made in integrating data to better serve students, policymakers, and the public.

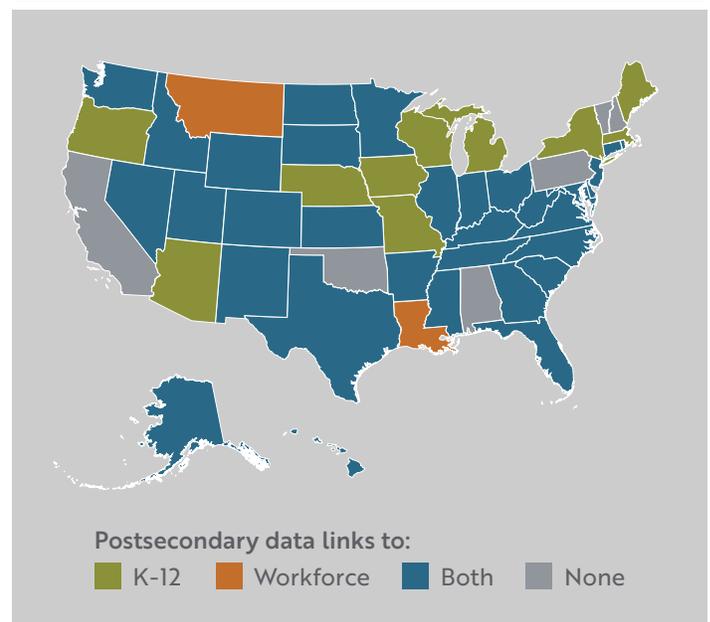
California's public education entities—K-12, the California Community Colleges (CCC), California State University (CSU), and University of California (UC)—maintain their own separate data systems and vary in how they collect and define data.<sup>1</sup> Private colleges each maintain their data separately and are not subject to any public reporting to the state on student outcomes. The California Student Aid Commission (CSAC) holds data on students' financial aid history. While each entity independently holds rich data, this segmented approach limits the public's knowledge of how the education system is performing, reduces decisionmakers' ability to make fact-based policy decisions, and prevents families from being informed consumers of higher education.

Transparency in public education is critical to making sound investments and serving students on the path from K-12 to college to the workforce. Currently, transparency in California ends when a student graduates from high school. **Forty-three states have or are developing statewide systems to link K-12 data with postsecondary data, and 33 link education data to workforce**

**data; California is on neither list, one of six states in the nation with no statewide, integrated education data system** (Figure 1).

While California has seen some efforts to share education data, progress on creating a robust and accessible information system has been limited. (See Appendix B for a history of legislation on data systems.) This brief offers specific policy solutions to move toward an integrated education and workforce data system, building on existing research on the shortcomings of the current data systems in California.

**Figure 1: States with statewide integrated education data systems**



Source: California Competes' aggregation of data from the State Higher Education Officers Association and the Education Commission of the States<sup>2</sup>

# Why California Needs an Integrated, Longitudinal Data System

For some students, the path from education to employment is a direct line from high school to college enrollment to a job. But for most students, the path is winding or broken. California currently allows K-12 and each of the three public higher education segments—independently from one another—to report their respective student outcomes. This approach does not reflect the complex and interconnected journey from education to the workforce. For example, because high school graduation data is not connected to college enrollment data, the state does not know how many California high school graduates enroll in college the following fall. The state also lacks a systematic understanding of student

mobility across higher education institutions—for example, students moving in- or out-of-state or attending multiple institutions. This mobility is the norm in California: over half of CSU graduates start out at a community college.<sup>3</sup>

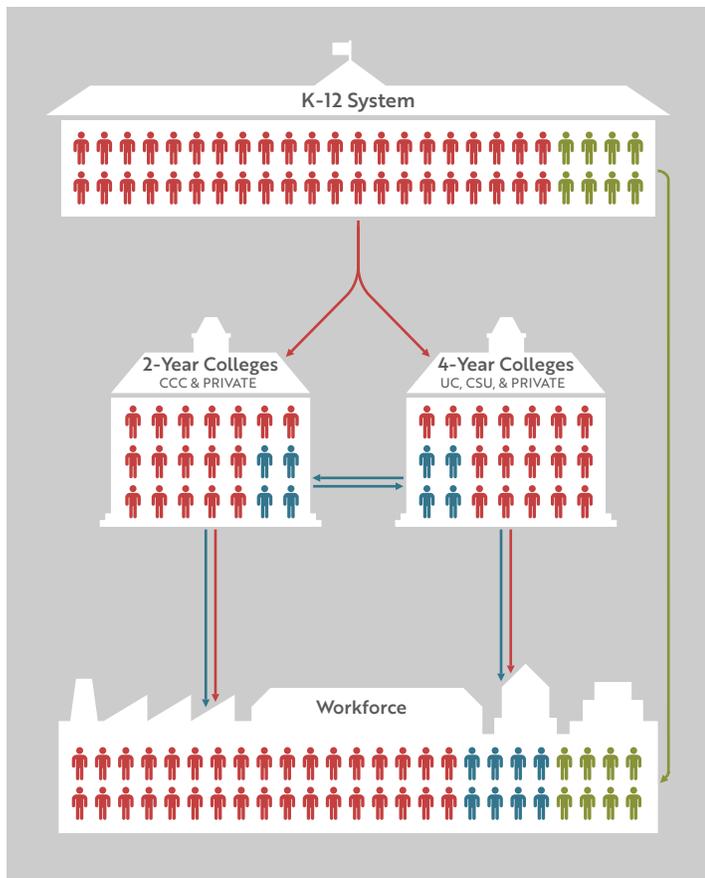
An integrated, longitudinal data system, on the other hand, illuminates students’ pathways from elementary school to college and into the workforce. Every transition point in the education-to-employment pipeline presents an opportunity to improve student outcomes by making state policy or institutional practice decisions based on accurate information (see Figure 2).

**Figure 2: What an integrated, longitudinal data system could say about California’s students**

Our state’s current education data system does not move with students across institutional boundaries. There are several pathways that students can take from education to workforce:

- K-12 to a single college to workforce
- K-12 to multiple colleges to workforce
- K-12 to workforce

With an integrated system, data would follow students across these transitions, illuminating where the pipeline is and is not working.



## How many California high school graduates enrolled in two-year and four-year colleges?

Understanding transitions from high school to college can help strengthen partnerships between K-12 and higher education, allowing for targeted outreach to the students and regions with the largest barriers to college enrollment.

## How many students who start at a two-year college go on to complete a certificate or degree, at the same college or after transferring?

Understanding transitions between colleges reveals common pathways for students and the success rates for students in each pathway. This information can help institutions collaborate to better serve transfer students, improve retention, and ease transitions.

## How do job placement rates and wages look for California high school graduates who do not attend college?

Understanding the transition from high school to workforce helps students and parents make informed decisions about their post-high school plans. It can also help the state make decisions about career and technical programs that need to be expanded or discontinued.

# Current Challenges to Integrating Data

Four key barriers in the current education data infrastructure must be addressed for the state to have better information about the education-to-employment pipeline.



## California lacks a higher education entity to coordinate data across segments

Without a coordinating agency for higher education in California, no authority exists to collect, maintain, and make decisions about higher education data across the three public segments of CCC, CSU, and UC, as well as private colleges. The California Postsecondary Education Commission (CPEC), which was defunded in 2011, once held this role, but the historical data CPEC housed was transferred to the California Community College Chancellor's Office (CCCCO) and is no longer updated.<sup>4</sup> The question of where an integrated, longitudinal data system should be housed is closely linked to the question of governance. To gain the trust and buy-in of all participants, the information needs to be stored in a way that is politically neutral, secure, and assures a balance of power among all segments. Absent any coordinating body, other collaborations, such as Cal-PASS Plus and the National Student Clearinghouse have tried to fill the gap. (See Q&A box on next page.) Institutions and K-12 districts can contribute their institutional data to either organization and

receive matched data showing subsequent college enrollments for the students they served. Because these private entities limit the data available to the public and not all entities choose to participate, the state is left with incomplete knowledge of higher education pathways and outcomes.



## Privacy laws and the segmentation of the K-12, higher education, and workforce systems limit data sharing

The Federal Education Records Privacy Act (FERPA) restricts institutions from sharing data on individual students without their consent but offers an exception that enables entities to share individual-level, identified student data for the purposes of auditing by a state or local education agency. Even when the legality of data sharing is clear, existing data-sharing collaboratives are voluntary and rely on individual agreements between entities. Without clear guidance that K-12 districts and higher education segments *can* and *must* share information, many institutions have opted not to participate in data-sharing initiatives. Even if legal and mandated, public concern over data privacy and security can make institutions reluctant to share data, particularly when they serve vulnerable populations such as undocumented students and workers. (See Q&A box below.)

### Q&A: Would an integrated, longitudinal education data system compromise data privacy and security?

Data security (practices that protect data from unwarranted uses) and privacy (the right of individuals to control who is able to see their data) are important concerns for any data system, especially one that involves transfers across multiple data owners. However, an integrated data system can be both secure and comprehensive if care is taken to adhere to best data safeguarding practices. Although FERPA is often referenced in data-sharing conversations, the law has not kept pace with technological innovations, leaving room for interpretation that can be complex for institutions to navigate.<sup>5</sup> Therefore, FERPA should be considered a minimum threshold to meet, not the gold standard. In response to recent data breaches, states have enacted myriad new laws to further protect individual privacy.

California's 2014 Student Online Personal Information Protection Act (see Appendix B for more detail) is seen as a model nationally for protecting students from having their data sold by online vendors. In addition to legal advances, there is a growing understanding of best practices in data storage, and organizations have begun to specialize in aiding institutions with data security and certifying institutions for best data security practices.<sup>6</sup> States that have successfully created secure integrated data systems have developed clear guidelines on the purpose, acceptable uses, and specific data elements that can be shared. Others have written into statute specific technical and physical safeguards for how data are to be stored, transferred, and accessed.<sup>7</sup> Please see Appendix A for further information and resources on data security.



## K-12, higher education, and workforce organizations use different data definitions and distinct student ID numbers

Segments, and often institutions within segments, currently each manage their own data collection and therefore create their own definitions of their data elements. Currently, demographic definitions vary; for example, whether Filipinos are aggregated along with Asians and/or Pacific Islanders. Fields of study are also not consistently defined; for example, the UC aggregates engineering and computer science majors in its external reporting, while the CSU reports them separately. Informal data partnerships do exist, but there is no entity that currently has the capacity or authority to standardize data. Working groups that have been tasked with creating common data definitions and standards across higher education segments have not sustained due to lack of funding and ownership.<sup>5</sup>

One challenge to linking data on students who move across segments is that the segments do not use a common student identification (ID) number. A common identifier is a code or number assigned to individual students to help manage their records. Currently, the California Department of Education (CDE) generates one ID number that is tied to a student from kindergarten through 12th grade, but each higher education segment then assigns its own separate ID number to students. Social security numbers are insufficient because not all

students have one and there are risks of negative ramifications for undocumented students. Furthermore, K-12 schools are prohibited from collecting social security numbers, and even at community colleges, more than 10% of students do not report a social security number.<sup>6</sup> While it is possible to have an integrated data system without common ID numbers, matches are less precise and require development of complex algorithms to match students based on other identifying information, such as names and birthdates. Sharing a common ID number can be more secure than sharing other identifying information to match students.<sup>7</sup> State legislation in 2008 required the higher education segments to collect students' K-12 identification numbers, but this legislation has not been funded or implemented.<sup>8</sup>



## California lacks a single portal for making education data available to the public

California's K-12 system, its three higher education segments, and individual private colleges each maintain separate portals for the public to access data.<sup>9</sup> These siloed data portals severely limit knowledge of trends in education pathways and also limit the ability to identify equity gaps. A unified portal to access integrated longitudinal data will improve transparency without adding burden for individual segments. Any public portal should provide data aggregated so that individual students cannot be identified.

### Q&A: What are our state's current data initiatives missing?

**Cal-PASS Plus** is a joint project of the CCCCCO; the Education Results Partnership, an independent nonprofit; and San Joaquin Delta College. Cal-PASS Plus aims to serve both as an infrastructure for linked high school and postsecondary data, as well as a technical assistance provider to help the data contributors utilize the information it collects. Cal-PASS Plus has recently added a feature that links to employment and earnings data, but only for community college graduates. To participate in Cal-PASS Plus, school districts and institutions sign MOUs that allow for data exchange. The majority of K-12 districts and many colleges in California have participated in Cal-PASS Plus to some degree, but not all are currently active. Cal-PASS Plus currently makes linked data available to participating institutions and very limited data for the general public.

**The National Student Clearinghouse (NSC)** is a non-profit company to which colleges voluntarily provide their enrollment data and K-12 school districts can pay a fee to receive linked data on their graduates' college pathways. As of this writing, approximately 98% of colleges nationwide, including public, private non-profits, and for-profits, contribute data to the NSC. The NSC includes basic data on student enrollments, majors, and degrees completed, but it does not include detailed course-taking and GPAs, limiting its utility to answer nuanced questions about student pathways. The NSC occasionally releases research reports, but data are not otherwise accessible to the public.

# A Policy Agenda for Education Data

Overhauling California’s education data system will require the state to address the four roadblocks outlined earlier. Implementing an integrated data system will require an investment from the state. The California Senate Committee on Appropriations analysis of SB 1224 (Glazer), which proposes a statewide data system and is being considered in the California Senate as of this writing, estimates a total cost “in the tens of millions” but cannot be precise because the bill does not specify the data to be collected.<sup>10</sup>

With these considerations in mind, we propose the following steps to put California on the path to an integrated data system that spans students’ entire educational careers into the workforce:



## Create a higher education coordinating entity.

The entity should be charged, among other roles, with collecting data and reporting on aggregate education inputs and outcomes, and:

- Be independent of the higher education segments and the California Department of Education (CDE) and be governed by a board of members from outside the institutions, appointed by the governor and legislature
- Have funding for staff to collect, analyze, and report data as well as develop policy options
- Be mandated to submit an annual report focused on integrated data indicators (rather than reporting on segments) and report to the legislature and governor, thus utilizing the auditing exception to data sharing under FERPA. This report would include, for example, statewide high school to college matriculation rates and degree production by major



## Develop clear guidance to ensure that data sharing across entities is safe, legal, and mandatory.

A statutory mandate to submit data will likely be required to ensure data sharing that provides clear guidance on privacy safeguards. Specifically, the state will need to:

- Engage the California Student Aid Commission (CSAC), each higher education segment, CDE, and private colleges in contributing the following data, as available, to the higher education coordinating entity on an individual level annually:
  - Student ID numbers
  - Demographic information (race/ethnicity, gender, SES)
  - Academic performance (e.g., high school GPA, course grades, SAT scores)
  - Institutional enrollment data
  - Participation in high school academic support programs
  - College major/program of study
  - Course enrollment and credits history
  - Federal, state, and institutional financial aid history
  - Graduation status
  - Loan repayment and default status
- Direct the coordinating entity to create a data-sharing agreement to match data with the Employment Development Department on:
  - Employment status
  - Wages
  - Workforce program participation
  - Disability claims
- Provide general fund dollars for the higher education segments and CDE to meet the reporting mandate
- Instruct the coordinating entity to implement careful security restrictions for how individual-level data are housed, limit access to individually identifiable data, and implement a regular process to audit data security practices



## Institute a process for the higher education segments, private colleges, and CDE to develop and adopt standardized data metrics and a common student ID number.

Data in their current state can be shared, but to ensure that individual-level data can be most informative and comparable across institutions and segments, all data contributors should collaborate toward developing and adhering to common data definitions and data collection standards. The state should ensure that this happens by:

- Creating an intersegmental advisory group which includes membership from K-12, CCC, UC, CSU, private colleges, and EDD to develop recommendations for data standardization
- Requiring higher education to adopt the K-12 state student ID and an algorithm for matching individuals in cases where the state ID number is not available
- Providing general fund dollars for the higher education segments and CDE to implement changes in their data collection to meet the new common data standards



## Develop an outlet for public reporting of integrated data.

The creation of a public portal to access linked data that has been aggregated to protect confidentiality is vital. This portal should be maintained by the state coordinating entity. To ensure that data are both openly accessible and responsibly used, the state should:

- Require the coordinating entity to create a public portal for sharing linked system data with the public
- Direct the intersegmental advisory group to institute a process for user controls that allows for secure access to detailed, de-identified data for vetted users, such as institutional researchers

## Conclusion

The blueprint in this brief presents policy solutions to address the state's need for integrated, longitudinal education data. Such a system will increase the transparency of education pathways in California and move education policy from an institution-centered to a student-centered approach. Our state needs to better understand how to optimize the student experience, build efficiencies, and scale interventions that succeed. The governance hurdles that stand in the way are complex, and it will take a significant initial investment of resources to make integrated data a reality. However, the technologies exist, there is precedent from other states as well as in California for

this type of integration, and cost efficiencies will be gained in the long run. An integrated data system is achievable without risking personal privacy or burdening institutions.

California is well-positioned to advance toward 21st century education data practices, illuminating our understanding of how California students navigate education pathways or falter along the way. This new system will empower students, families, educators, and policymakers with the knowledge necessary to improve higher education in California and increase prosperity and social mobility for individuals, their families, and communities.

## Q&A: What are other states doing with educational data?

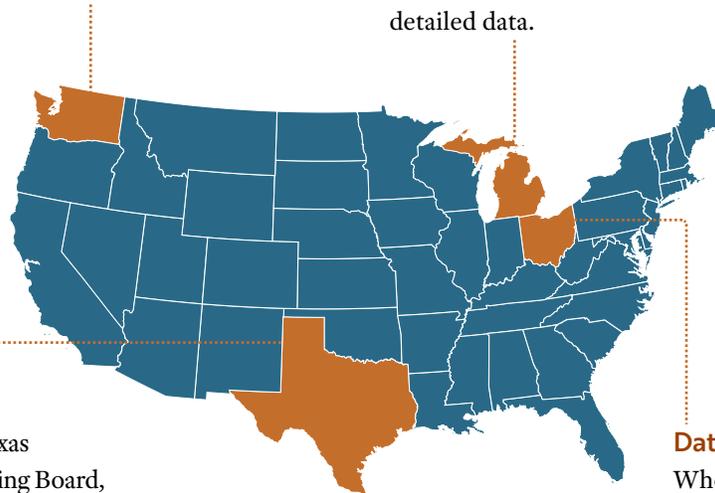
Because California would be among the last states to arrive at an integrated education and workforce data system, there is plenty of opportunity to learn from other states. The following highlights how other states have successfully navigated the four policy areas outlined in this section.

### Standardizing Metrics Across Entities:

Washington state developed a longitudinal data system called the Education Research and Data Center (ERDC) by legislative mandate in 2007. The ERDC is governed by three technical advisory committees, one of which, the Data Stewards Committee, consists of research staff from institutions that contribute data and is charged with maintaining consistent data definitions and making recommendations for data collection.<sup>11</sup>

### Making Linked Data Publicly Available:

Michigan, the only state other than California that does not have a higher education coordinating entity, created a data dashboard that makes data on K-12, postsecondary, and transitions between the two available to the public. The dashboard, called **MI School Data**, is housed by the Michigan Department of Education. This portal has some data available for the public to query, as well as a secure section where authorized users can log in to obtain more detailed data.



**Data Governance:** The Texas Higher Education Coordinating Board, created in 1965, is responsible for leadership, planning, and auditing the state's higher education system. The board consists of nine members who are appointed by the governor and represent diverse sectors of the business community. The board staffs several advisory committees and departments that work on a wide array of areas of coordination, including auditing and compliance, financial planning, and college readiness. The Board oversees and houses the state's integrated higher education data system.

**Data Privacy and Security:** When the Ohio legislature created an integrated longitudinal system in 2010, it included strict restrictions on what data were to be included and what specific positions in educational agencies could access the data. Ohio restricts the inclusion of student names, relying instead on a common ID number, to reduce the risk of students being identified. The legislature also mandated an annual report on who requested access to data and how it was used.<sup>12</sup>

Please see Appendix B for references to more state statutes related to data systems.

# Notes

1. Moore, C., Reeves Braco, K., and Nodine, T. (2017, August). *California's maze of student information: Education data systems leave critical questions unanswered*. Education Insights Center. [http://www.mtsac.edu/president/cabinet-notes/2017-18/Californias\\_Maze\\_of\\_Student\\_Information\\_August\\_2017.pdf](http://www.mtsac.edu/president/cabinet-notes/2017-18/Californias_Maze_of_Student_Information_August_2017.pdf)
2. Armstrong, J., and Whitfield, C. (2016). *Strong Foundations 2016: The state of state postsecondary data systems*. State Higher Education Officers Organization. [http://www.sheeo.org/sites/default/files/publications/SHEEO\\_StrongFoundations2016\\_FINAL.pdf](http://www.sheeo.org/sites/default/files/publications/SHEEO_StrongFoundations2016_FINAL.pdf)  
Education Commission of the States (2016, October). *State longitudinal data systems - All state profiles*. <http://ecs.force.com/mbdata/mbprofallrtan?Rep=SLDS16STA>
3. Friedmann, E. (2017, June). *Building intersegmental partnerships*. Policy Analysis for California Education. <https://edpolicyinca.org/sites/default/files/Building%20Intersegmental%20Partnerships.pdf>
4. California Senate Appropriations Committee (2011, May 24). SB 1138 (Liu) – *California Postsecondary Education Commission (CPEC) data*. [http://leginfo.legislature.ca.gov/faces/billAnalysisClient.xhtml?bill\\_id=201120120SB1138#](http://leginfo.legislature.ca.gov/faces/billAnalysisClient.xhtml?bill_id=201120120SB1138#)
5. Grama, J. L. (2016, May). *Understanding information security and privacy in postsecondary education data Systems*. Institute for Higher Education Policy. [http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/information\\_security\\_and\\_privacy.pdf](http://www.ihep.org/sites/default/files/uploads/postsecdata/docs/resources/information_security_and_privacy.pdf)
6. Mathewson, T. G. (2017, October 18). Who is keeping student data safe in the era of digital learning? *The Hechinger Report*. <http://hechingerreport.org/keeping-student-data-safe-era-digital-learning/>
7. EducationCounsel, LLC (2014, March 3). *Key elements for strengthening state laws and policies pertaining to student data use, privacy, and security: Guidance for state policymakers*. <https://2pido73em67o3eytaq1cp8au-wpengine.netdna-ssl.com/wp-content/uploads/2016/03/Ed-Counsel-Guidance-on-State-Student-Privacy-and-Security-Policies.pdf>
8. Warren, P., and Hough, H., (2013, August). *Increasing the usefulness of California's education data*. Public Policy Institute of California. [http://www.ppi.ppic.org/content/pubs/report/R\\_813PWR.pdf](http://www.ppi.ppic.org/content/pubs/report/R_813PWR.pdf)
9. California Community Colleges Chancellor's Office (2017). *Analysis: Missing Social Security Numbers (SSN) in the California Community Colleges (CCC) management information system*. [http://extranet.cccco.edu/Portals/1/TRIS/Research/Accountability/ARCC2\\_0/MissingSSNs2017.pdf](http://extranet.cccco.edu/Portals/1/TRIS/Research/Accountability/ARCC2_0/MissingSSNs2017.pdf)
10. Ohio Department of Education (2014, December). *Data privacy report*. [https://education.ohio.gov/getattachment/Topics/Data/EMIS/Data\\_Priv\\_Rpt.pdf.aspx](https://education.ohio.gov/getattachment/Topics/Data/EMIS/Data_Priv_Rpt.pdf.aspx)
11. Perry, P., Furgiuele, C., and Garcia, P. (2015). *K20W intersegmental data collaborative*. Presentation to the California Association for Institutional Research. [https://cair.org/wp-content/uploads/sites/474/2015/07/5\\_Perry.pdf](https://cair.org/wp-content/uploads/sites/474/2015/07/5_Perry.pdf)
12. Warren and Hough, 2013.
13. California Senate Appropriations Committee (2018, April 30). SB 1224 (Glazer) – *Statewide longitudinal education and workforce data system*. [https://leginfo.legislature.ca.gov/faces/billAnalysisClient.xhtml?bill\\_id=201720180SB1224#](https://leginfo.legislature.ca.gov/faces/billAnalysisClient.xhtml?bill_id=201720180SB1224#)
14. Washington Office of Financial Management (2015). *Education Research and Data Center: Data governance*. <https://erdc.wa.gov/about-us/data-governance>
15. Ohio Department of Education, 2014
16. California EDGE Coalition (2016, June). *Data to serve policy, programs, and people: Reinventing California's education and workforce data systems*. [http://californiaedgcampaign.org/wp-content/uploads/2016/06/CA\\_EDGE\\_DatatoServe16-2-FINAL.pdf](http://californiaedgcampaign.org/wp-content/uploads/2016/06/CA_EDGE_DatatoServe16-2-FINAL.pdf)

*This publication was produced with support from the Bill & Melinda Gates Foundation, along with support from the College Futures Foundation, Lumina Foundation, and James Irvine Foundation.*



# Appendix A.

## Experts and Publications on Integrated Data Systems

Several organizations have recently documented challenges with California’s education data infrastructure and the potential benefits of a better system; their research informed this brief. Below is a summary of relevant publications.

For more information on...	Please see...	By...
<b>Data Governance</b>		
California history of higher education data policy	<a href="#">California education policy, student data, and the quest to improve student progress</a>	Education Insights Center
Federal policy on higher education data	<a href="#">Better data on the horizon: An analysis of evolving student-level data legislation</a>  <a href="#">Envisioning the national postsecondary data infrastructure in the 21st Century</a>	Institution for Higher Education Policy
Examples of how other states have built their integrated data infrastructure	<a href="#">State Longitudinal Data Systems – All State Profiles</a>	Education Commission of the States
Workforce data policy and linkages to higher education	<a href="#">Data to Serve Policy, Programs, and People: Reinventing California’s Education and Workforce Data Systems</a>	EDGE Coalition
<b>Data Privacy and Security</b>		
Data security best practices and sample legislative language	<a href="#">Key Elements for Strengthening State Laws and Policies Pertaining to Student Data Use, Privacy, and Security: Guidance for State Policymakers</a>	Education Counsel
<b>Standardizing Metrics Across Entities</b>		
Data system considerations for California	<a href="#">Increasing the Usefulness of California’s Education Data</a>	Public Policy Institute of California (PPIC)
<b>Making Linked Data Publicly Available</b>		
Equity and using data for community advocacy	Forthcoming	Education Trust West

## Appendix B. Summary of Relevant Prior and Current Legislation

Bill	Key Provisions	Status
<b>California</b>		
SB 1224 (2018)	Requires the CDE, CCC, CSU, and requests the UC to establish a data system that tracks students longitudinally	Currently being debated. See <a href="#">here</a> for status.
AB 1936 (2018)	Establishes a state higher education coordinating entity and requires it to establish and maintain a statewide longitudinal data system	Currently being debated. See <a href="#">here</a> for status.
SB 42 (2016)	Would have established a state higher education coordinating entity (like the current AB 1936)	Vetoed by Governor Brown.
SB 1177 (2014) (Student Online Personal Information Protection Act)	Restricts sharing or selling of student information by any online service	Chaptered.
AB 2148 (2014)	Encourages data integration between career and technical education programs and workforce entities by creating an interagency workforce metrics dashboard	Chaptered, but the data requirements have not been implemented. <sup>16</sup>
SB 1136 (2011)	Would have created a data management system to take over the CPEC's data role	Held in Appropriations.
<b>Federal</b>		
College Transparency Act	Establishes a national, student-level postsecondary database  Requires postsecondary data to be matched to data from other agencies, including the Department of Treasury, Veterans Affairs, and Social Security	Currently being debated. See <a href="#">here</a> for status.
Student Right to Know Before You Go Bill (H.R. 4779/S. 1195)	Requires colleges that receive federal aid to report student data to the Department of Education for the purposes of public reporting on student outcomes.  Includes specific privacy protections, including forbidding data from being sold or being used to take action against students.	Currently being debated. See <a href="#">here</a> for status. This bill has been introduced every session since 2012 and has not passed.
PROSPER Act (HR 4508)	Reauthorizes the Higher Education Act and includes a provision to maintain a ban on collecting student-level data	Currently being debated. See <a href="#">here</a> for status.