



CALIFORNIA Data System

Technology & Security Sub-Committee Background Paper 1: Frameworks for Structuring Data in a Statewide Data System

Introduction

In 2019, California enacted the Cradle-to-Career Data System Act (Act), which calls for the establishment of a state longitudinal data system to link existing education, social services, and workforce information.¹ The Act also lays out a long-term vision for putting these data to work to improve education, social, and employment outcomes for all Californians, with a focus on identifying opportunity disparities in these areas.

The legislation articulated the scope of an 18-month planning process for a linked longitudinal data system. The process will be shaped by a Workgroup that consists of the partner entities named in the California Cradle-to-Career Data System Act.² Suggestions from this workgroup will be used to inform a report to the legislature and shape the state data system designs approved by the Governor's Office. Because the legislation describes a number of highly technical topics that must be addressed as part of the legislative report, five sub-committees were created that include representatives from the partner entities and other experts. The Technology & Security Sub-Committee

1 Read the California Cradle-to-Career Data System Act at:
https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=EDC&division=1.&title=1.&part=7.&chapter=8.5.&article=

2 The partner entities include the Association of Independent California Colleges and Universities, Bureau for Private Postsecondary Education, California Community Colleges, California Department of Education, California Department of Social Services, California Department of Technology, California Health and Human Services Agency, California School Information Services, California State University, California Student Aid Commission, Commission on Teacher Credentialing, Employment Development Department, Labor and Workforce Development Agency, State Board of Education, and University of California.

will help to develop technology specification requirements to address data structures and security considerations.

This brief is intended to provide a discussion framework for members of the Technology & Security Sub-Committee. It includes background information on the authorizing legislation and a summary of the priorities established by the Workgroup for phase one of the data system. Next, it defines common data structures used to construct state systems including centralized, federated, and hybrid systems, with special consideration for concepts like data lakes and real-time data. In addition, examples from five states are provided to clarify various models for structuring data. The brief concludes with framing questions that the subcommittee will consider at their first meeting, in order to recommend which data structures should be used for the first phase of state data system development.

The California Cradle-to-Career Data System Act

In 2019, California enacted the Cradle-to-Career Data System Act, which outlines the scope of an 18-month planning process, allocates \$2 million to support that process, and earmarks an initial \$10 million toward the development of a state data system.

The Act also lays out a long-term vision for putting data to work to improve outcomes for all Californians, with a focus on identifying disparities in opportunities. By securely linking data that schools, colleges, social service agencies, financial aid providers, and employers already collect, the data system will

- enable users to identify the types of supports that help more students learn, stay in school, prepare for college, graduate, and secure a job;
- provide information that teachers, parents, advisors, and students can use to identify opportunities and make decisions;
- help agencies plan for and improve education, workforce, and health and human services programs; and
- support research to ensure policy effectively supports individuals from birth through career.

Recognizing that the data system will need to be built in phases, the California Cradle-to-Career Data System Act lays out several priorities:

- **Linking existing information in the system.** The first data sets to be linked should be existing K–12 and college data sets, followed by employment and earnings

data, early childhood education information, and social services information, although this order can be amended.

- **Guaranteeing privacy and security.** The system cannot be built until clear guidelines and legal agreements have been established to ensure that information will be securely gathered and stored in compliance with federal and state laws and in accordance with privacy best practices, and that the identity of sensitive populations will be protected.
- **Providing information for students, families, and educators.** The system will include an interface for sharing information with teachers, parents, advisors, and students.
- **Facilitating analyses for researchers and policymakers.** The system will link data between agencies to help answer foundational questions about the impact of state policies and investments.
- **Assuring quality.** The legislation addresses the need to improve the quality and reliability of education information, both within and between agencies and other entities providing data.

Priorities for Phase One of the Data System

In the first meeting of the Workgroup, the partner entities recommended that the California Cradle-to-Career data system should be an ecosystem that allows for various tools, processes, and resources to be developed under a governance structure. In its first phase, the state should build a P20W data set that includes early care, K–12, postsecondary, financial aid, and employment information. This data set should be used to create dashboards that provide useful information for practitioners and the public, as well as query tools that allow for more nuanced analyses. The P20W data system should be paired with a clearly defined process to link additional data points as needed to answer inquiries, including requests from outside entities such as researchers, policymakers, and regional partnerships, as well as to foster the secure exchange of information between partner entities. Finally, the Workgroup recommended that the state develop tools that provide information directly to individuals or allow teachers and counselors to better understand the needs of the people they serve. Possible options will be examined at the February 2020 Workgroup meeting, based on tools that have been built in other states or developed in California but not implemented statewide, such as alerting students about the social service and health benefits that they are eligible for, informing educators about services that a student is receiving, or creating an e-transcript service to support college and financial aid applications.

Data Structures for State Data Sets

As a first step in creating the P20W data set, the Technology & Security Subcommittee is tasked with considering the best data structure to produce the types of information that have been prioritized for phase one of the data system, while ensuring that the structure does not preclude future desired data use cases.

In many analyses of possible data structures, federated and centralized systems are presented as two contrasting options (Moore & Bracco, 2018). Federated data systems temporarily link information from participating agencies to address a specified purpose, while centralized data systems gather information from the agencies and store it in one place (Levantoff, n.d.). However, all states currently maintain either a centralized or a hybrid data system.

Due to federal requirements under the Every Student Succeeds Act (ESSA), which mandates that historical information on student outcomes be maintained, states that had federated systems have created centralized repositories that can store this information. In addition to these centralized structures, some states maintain federated structures to link data for specific purposes such as developing predictive models for postsecondary outcomes based on risk factors or following cohorts of high school graduates to examine their economic mobility (Institute of Education Sciences, 2019).

While most states employ a mixture of approaches, federated and centralized systems have divergent requirements and offer different trade-offs, which can be useful when determining when to select a model for a specific use case. For example, federated and centralized models involve different levels of agency control, complexity of inter-agency agreements, recency of the data, requirements for data alignment, and costs.

Federated Models

Federated systems query and retrieve information from a variety of existing siloed data sets but do not keep the results after the analysis has been conducted. They can integrate multiple types of data including transactional systems (such as CCC Apply), existing data warehouses (such as CALPADS), or live production systems (such as the California Colleges Guidance Initiative).

Federated approaches provide a perception of additional security and control from participating agencies. They can implement rules that allow certain queries to process

information and deny others that don't meet pre-set criteria. They also provide a higher degree of privacy protections because information is not stored in a centralized location. However, given that federated systems generally don't allow storage over time, they don't maintain snapshots of information that help facilitate longitudinal analyses.

Figure 1: Federated Model – Pros & Cons

Pro	Con
Perception of source agency control	Capabilities limited by source agency systems and rely on source data being available
Legal authority and access issues are easier to achieve because queries are constrained by access rules	If one of the source agencies doesn't support the query, it cannot be implemented—an issue that compounds as more source agencies are included in the data system
Data likely to be more recent	Queries and data element mapping must be developed for each new request
Stronger privacy protections can be implemented because information isn't kept in one place	Longitudinal analyses are more difficult to do because there are no historical snapshots of data
Cheaper to build	Requires more resources from partner entities to address each query
Several state models to learn from	States using federated systems struggle to release data to the public or outside sources

Centralized Models

Centralized data systems house a set of unitary student records with pre-determined data elements that are regularly refreshed by partner entities. Centralized systems can include multiple types of data, so long as underlying definitions have been aligned. The majority of state data systems are centralized, and in almost all cases data are stored in a de-identified format.

Centralized approaches provide a perception of greater access because information is already compiled in one place and not subject to approval before data can be combined. Centralized systems are also responsive to the variable nature of education data because information from a specific year can be analyzed in the context of requirements at that time, such as taking into account adjustments to cut scores on assessments or legislative changes to graduation requirements. They also provide a built-in focus on alignment because specifications must be created for how data elements should be constructed by each agency in order to harmonize results. However, given that centralized systems use scheduled refreshes of the underlying data, they tend to have information that is less current.

Figure 2: Centralized Model – Pros & Cons

Pro	Con
Perception of greater access	Data must be extracted, transformed, and loaded to fit into the data warehouse specifications
Data governance policies clarify allowable uses and which elements should be stored centrally for those purposes	Governance of a centralized warehouse is more complicated
Data alignment is easier to monitor because documentation tracks how data elements and metrics change over time	Data are generally not real-time
Many education and technology vendors have very mature tools and procedures to load, secure, and display combined data sets	Privacy and security rules may be more complex given that data are stored in one place
After initial criteria are established, less work is required of the partner entities	More expensive to build
Several state models to learn from	Legal issues with access and hosting can derail an otherwise simple technical architecture

Understanding Data Lakes

Recent shifts in technology have made it possible to process information more quickly and store vast amounts of data in the cloud. For example, a data lake is a virtual file storage system that allows each data contributor to provide information in its native format. Instead of requiring each entity to submit information in a standardized manner, such as converting the indicator for females to “F,” contributors can upload whatever variable they use, such as “female,” “woman,” “girl,” or “1.” Data lakes may use mapping tables to track which variable is used by each entity so that the concepts can be compared, or they may require contributors to tag all data points to indicate which variable they represent. From the data lake, data warehouses can be constructed for analytical purposes.

No other state has used a data lake to create a longitudinal data system. However, the California State University Chancellor's Office has created a data lake and is currently working with each campus to validate information that has been produced from this centralized storage hub. To create the data lake, the Chancellor's Office leveraged an existing file backup capability, which captured full copies of each campus' data each day. Each of those daily copies is stored in Amazon Web Services (AWS), so that information can be extracted for any specific point in time. Using AWS RedShift, data are sorted into tables, called curated student collections, which are organized around key concepts such as Students by Term or Classes by Section. A role-based security system ensures that only authorized individuals can access the collections to view specific data elements or conduct analytical studies. Over time, this data set could be used to support research or build dashboards, both at the college and system level.

Approaches to Data Infrastructure

As the subcommittee grapples with questions about how to structure information for the Cradle-to-Career data system, it may be helpful to consider approaches implemented in other states or on a smaller scale in California. The examples below are intended to

provide a general overview of different ways to combine federated and centralized models and provide access to information.³

Virginia

Virginia developed its state data system using a federated model, allowing information from K–12, public postsecondary, employment, and healthcare agencies to be linked together on an as-needed basis. Participating agencies established governance policies and then built a data system that operationalized those agreements. The state data system includes a research portal that supports the process of vetting requests to access data for discrete purposes.

To link records for research purposes, each contributing agency replaces personally identifiable information such as names, birthdates, and testing identifiers with a randomly generated identifier. Once the data are moved into the state system, the agency-generated identifier is replaced by a new random identifier. The state system identifiers are destroyed after each use, to ensure there is no way to re-identify individual student information.

Data required to document postsecondary outcomes for K–12 students are generated using a separate centralized data warehouse. Reports, which can be created at the school or county level, are posted on the K–12 agency website and can be downloaded either as PDF or Excel files that contain aggregate information. The reports document longitudinal outcomes such as the percentage of students who earned one year's worth of college credit within two years of last enrolling in high school.⁴

Nevada

Similar to Virginia, Nevada's data system began as a federated approach that links K–12, public postsecondary, and employment information, but now has a hybrid approach and stores some information in a centralized data warehouse. Nevada contracted with the same vendor that produced Virginia's state data system to create

³ Members of the subcommittee and representatives from the partner entities will be invited to a professional development opportunity on March 9, 2020, where representatives from states with well-established systems will provide information on their policies and procedures related to data matching.

⁴ For more information about the Virginia system, visit <http://vlds.virginia.gov/>

its data infrastructure. Student records are matched using an algorithm, after which personally identifiable information is removed so that it is not imported into the state data system. Researchers can submit data requests to receive de-identified records. Data that are pulled for research requests use one-time identifiers, so that information cannot be re-identified.

Centralized data are used to create a series of dashboards that allow users to examine topics such as: remediation rates for postsecondary students compared to their high school ACT scores; postsecondary grades in English, math, or science; whether students earned a postsecondary award; and earnings of college graduates over a variety of time periods. Information can be sorted by K–12 district or viewed statewide.⁵

Minnesota

Minnesota's state data system is a centralized warehouse that includes K–12, postsecondary, adult education, workforce training, and employment information. A second centralized data system combines early care, K–12, social services, and health data. As is the case in Virginia and Nevada, after personally identifiable information is used to match records for each of the two longitudinal systems, sensitive information is removed and not imported into the data warehouses. The state is in the midst of transitioning to a new linking process that will allow information on parents and children to be connected, to better understand whether social service and education interventions yield multi-generational benefits. Minnesota has a request process by which approved researchers can receive files with de-identified student information.

For the early care longitudinal system, charts show a variety of data points such as how many kindergarteners participated in early care, whether they received economic and food assistance, academic outcomes in grade 3, and the relationship between maternal traits and education measures. In the P20W system, information is made available through charts that are specific to K–12 and postsecondary institutions. For example, users can see how many K–12 students enrolled in college, took remedial math and English courses, and how many credits they took in their first year of college. For college students, longitudinal information is available on graduation, employment, industry of employment, and average earnings. Users can filter results by education institution, region, graduation year, race, gender, and student characteristics such as

⁵ For more information about the Nevada system, visit <http://npwr.nv.gov/>

whether they were English language learners or participated in career and technical education programs.⁶

Kentucky

Kentucky has a centralized state data system that brings together K–12, postsecondary, adult education, workforce, employment, and teacher certification data. As is the case in the three other states, records are linked using a matching algorithm, a unique identifier is assigned, and personally identifiable information is removed. A virtual warehouse stores de-identified data that can be queried for dashboards, research studies, and information requests. Researchers can request aggregate data sets for which no legal agreement is needed or enter into a data use agreement to conduct approved studies using de-identified unitary data.

Publicly available information is provided through web-based infographics that can be tailored to display results for specific populations of students and education institutions. For example, K–12 information can be sorted by race, gender, special education status, and free and reduced lunch status and shown at the school and district level. The infographics combine numerous data points in one view. Sample visuals include how kindergarten readiness relates to math and reading scores in grade 3, which postsecondary institutions K–12 students enrolled in, and post-transfer and post-college employment outcomes by major.⁷

California

Several California organizations and projects have linked together a significant amount of data from different agencies or integrated education and social service information. These examples can shed light on how to structure a state data system or could potentially be scaled to support a broader scope of information.

Cal-PASS Plus is a voluntary clearinghouse of longitudinal data that combines student records from K–12, community college, California State University, University of California, and the Employment Development Department Unemployment Insurance wage file. It hosts a number of dashboards, including the LaunchBoard system and public K–12 achievement data. Cal-PASS Plus is funded by the California Community

⁶ For more information about the Minnesota systems, visit <http://sleds.mn.gov/#>

⁷ For more information about the Kentucky system, visit <https://kcews.ky.gov/>

Colleges Chancellor's Office through a grant to San Joaquin Delta College. The data system is administered by a nonprofit called Educational Results Partnership (ERP).

Information on all community college students and their employment outcomes is provided by the Chancellor's Office, whereas information on other types of students must be uploaded by each local education agency or college. All information is transferred via secure servers. While Cal-PASS Plus has 1,500 member institutions that provide data on the majority of students in California, the voluntary nature of participation means that some schools and some years of data are missing.

Cal-PASS Plus manages data using the same process found in most centralized state data systems. ERP loads and encrypts raw data and matches records to existing data using an algorithm. Once matching is complete, personally identifiable information is removed and the information is moved to a data store. There, ERP loads the information into a schema and transforms it to normalize information into categories such as students, courses, and awards. The files are examined to determine whether information is missing and if it is in the appropriate format. From the data store, ERP builds temporary data sets that are specific to individual projects. During the process of building data tools or conducting analyses, information is stored on servers. Once data sets are finalized, aggregate information is stored in the cloud on both Amazon Web Services and Microsoft Azure. ERP also uses an Application Programming Interface (API) to provide recommendations to community colleges regarding whether students would benefit from receiving support in first-year math and English courses, based on their high school GPA and course-taking patterns.⁸

Cal-PASS Plus has more education data elements than would be required for the P20W data set and does not contain early care or financial aid information, but it has produced dashboards from the underlying data set that include K–12, postsecondary, and workforce information. This structure could be scaled and modified to create the P20W data set and provide additional information for specific requests. It is not currently structured in a way that would produce tools for individuals or practitioners.

The Children's Data Network (CDN), which is affiliated with the University of Southern California, is partnering with the California Health and Human Services Agency (CHHS) to create a cross-department, centralized data sharing portal called the Research Data Hub. Using grant funding, the project builds upon work that the CDN has already

⁸ Interview with Anthony Dalton, January 14, 2020.

been doing to match records across various CHHS departments and programs to identify whether individuals are receiving multiple services, such as determining whether an individual receiving health insurance through Medi-Cal is also receiving food assistance through CalFresh. The Research Data Hub will use historical, de-identified data sets provided by the individual CHHS departments to advance knowledge and support the delivery of person-centered health and human services. In addition, it has been designed to link individual-level data across CHHS programs for more operational use cases.

The Research Data Hub expands upon the legal and governance processes embedded in the Inter-Agency Data Exchange Agreement (IDEA), a master CHHS agreement that is currently being revised and restructured by the Governor's Office to be applied across all executive branch entities. Each CHHS department and program can contribute person- and program-level administrative data to secure servers at the University of Southern California. There, data are separated into two types: de-identified information that can be used for research and personally identifiable information that is used to link records across programs. The CDN identifies individuals who are found across multiple data sets via a machine learning algorithm. For example, if Individual A from Medi-Cal has similar personally identifiable information to Individual B from CalFresh, the algorithm assigns a high probability to this pair of records, indicating a high level of confidence that they reflect the same individual. The CDN generates an encrypted linkage key for each pairwise person match. For instance, one linkage key is created to connect an individual between Medi-Cal and CalFresh records, and a separate linkage key would be generated for the same individual between Medi-Cal and CalWORKs data. Once the encrypted linkage keys have been created, they are transferred to the Research Data Hub along with de-identified information approved by CHHS departments.

The Research Data Hub will be a cloud-based environment where de-identified, individual-level information will be stored in structured data tables. Access to data sets will be restricted based on roles that are set by CHHS departments, which will have the ability to specify row- and column-level access for each data set. Approved CHHS staff or researchers will have remote access to the data and use analytic tools that are hosted within the Research Data Hub to generate cross-program analyses. Longer-term, the Research Data Hub will develop protocols that permit approved external research

partners to access curated data sets and statistical resources in this analytic environment.⁹

The Research Data Hub only has health and human service information, so all of the required data sources for the P20W data set or the requests for additional information would need to be added to the existing data set. Some visualizations on the CHHS Open Data Portal were created using information from the CDN matching process, which could provide valuable lessons for developing dashboards for the P20W system. The match process created by the CDN could be used to help integrate health and human service data into tools for individuals and practitioners.

California Colleges Guidance Initiative (CCGI) houses, audits, and transmits data to support decisions regarding admissions, financial aid, and course placement. It also provides tools that help with college and career planning through the website CaliforniaColleges.edu. CCGI receives funding from the legislature, foundations, and member institutions, and is managed under the fiscal agency of the Foundation for California Community Colleges.

CCGI's approach to managing data is significantly different from most state data systems because it interacts directly with individual education entities rather than state agencies. CCGI secures data on student demographics and course-taking directly from K–12 providers using an automated secure file exchange and stores it in Amazon Web Services. Currently, 81 districts that serve 520,000 students participate in CCGI. At the point that a student applies to a community college or California State University or for state financial aid, CCGI captures the student's postsecondary state identifier and adds it to the K–12 record. By holding a limited number of data points, CCGI minimizes risk while also creating a key that can be used in the future to link additional information.

The way that data are shared varies by postsecondary partner and by how a student applies to college. If a student launches a college application from their individual account at CaliforniaColleges.edu, a unique token that is associated with the student is sent to CCC Apply or Cal State Apply. When the student submits the application, CCGI adds the student's K–12 identifier. In the case of CCC Apply, if a student does not apply through the CCGI portal, the college application system submits the student's name, date of birth, and high school to the California Department of Education, where an

⁹ Interview with Emily Putnam-Hornstein, January 28, 2020.

algorithm matches the student and provides the K–12 State Student Identifier (SSID). Data for community colleges is transmitted nearly instantaneously. CCGI is alerted once the SSID has been matched, and if the student is from a district that participates in CCGI, the student's transcript is transmitted to the California Community Colleges Chancellor's Office. For CSU, which shares application status information daily, students have the option of allowing CCGI to automatically populate their coursework into the application using real-time high school information.

CCGI uses a similar process to provide data for financial aid applications. If a student launches an application to either FAFSA or the Dream Act from CaliforniaColleges.edu, the same match process is used to secure the student's K–12 identifier and provide it to the California Student Aid Commission. This process is implemented weekly. CCGI is currently working with the University of California to integrate with its application as well.¹⁰

CCGI does not include early care, postsecondary, financial aid, or workforce data needed for the P20W data set or the requests for additional information. It also collects data directly from local education agencies, rather than from state agencies. The Workgroup determined that it would be preferable to use agency data for phase one of the California data system. However, the token system could help provide stronger matches for education and financial aid data. CCGI could become one of the data tools for individuals and practitioners if it were scaled statewide to provide e-transcript services for college applications, generate course recommendations for community college math and English courses, and streamline financial aid applications.

The Silicon Valley Regional Data Trust (SVRDT) is an initiative of the Santa Clara County Office of Education, three Bay Area counties, and the University of California, Santa Cruz. With support from federal and private grants, SVRDT provides a Secure Data Environment, which will enable K–12 public schools, juvenile justice, and county health and human service agencies to share data to coordinate case management, personalize and integrate services, inform public policy, and develop a research-practice partnership with the University of California, Santa Cruz with the goal of improving student safety and educational outcomes. Having successfully implemented a prototype, SVRDT is in the midst of deploying its first county-level instance of the SVRDT data set.

¹⁰ Interview with Ben Baird, January 13, 2020.

The data structure for SVRDT reflects an underlying set of agreements between the partners about who should use the data and for what purpose. Having worked closely with both leadership and front-line practitioners over a series of years, SVRDT staff engaged practitioners from the participating agencies to determine the specific types of data that would be most actionable, including analysis by the SVRDT's Legal Working Group to ensure that each data point could be shared in the context of state and federal regulations. The “up front” data elements focus on such concepts as whether a student was participating in a social service program, on what dates, and for how long. SVRDT also created a common consent form that will be used across all education institutions and county agencies, paired with a consent management system, to ensure that appropriate permissions are in place before data are shared. Access to data is role based, with the partner entities determining the roles.

SVRDT's Secure Data Environment is a multi-agency transaction processing environment that supports active inquiries on children when they are being served by more than one agency. Information is drawn from a regional K–12 data warehouse, as well as county and state data systems. Rather than create a unique identifier, an SVRDT internal index records the identifier from each agency that provided services for a specific child. Because the source data resides in systems that are refreshed on a daily basis, every time new data are integrated or a query is run, the index is updated. SVRDT determines whether a child is already in the data set by matching specific elements such as names, dates of birth, and addresses. This personally identifiable information is encrypted and there are no direct access services to where the index is stored.¹¹

Similar to CCGI, SVRDT does not include many of the types of data needed for the P20W data set or the requests for additional information and pulls some information from regional sources rather than state systems. However, SVRDT provides a model for how to link education and non-education records, and for generating actionable data for practitioners, which could be used to design dashboards and tools that provide information to teachers, counselors, and service providers.

There are several other scaled data projects in California, such as the CORE Districts or Linked Learning Analytics. These centralized data warehouses, which focus on information from a single segment, can also serve as resources regarding how to organize, display, and share data in a manner that is useful for practitioners.

¹¹ Interview with Marcy Lauck, January 27, 2020.

Unpacking the Concept of Real-Time Data

Many partner entities and advisory group members expressed a desire for the California data system to provide “real-time data.” Whether data can be instantaneous relies on two factors in the context of state data systems: how frequently information is submitted to the linked data system and, given that these data systems generally pull information from state agencies, how frequently individual institutions submit data to those agencies.

The frequency of data provided by partner entities may be close to real time, such as nightly or weekly uploads provided through an API or a transfer of files to a secure server. However, if education institutions only submit information to a state agency once a term or once a year, data on student progress and outcomes will be lagged. For example, for federal reporting, employment and earnings figures are measured for students six months and one year after they left the education system. This means that metrics can only be calculated once enrollment data demonstrates that the student did not resume coursework for a full year. If colleges don’t submit enrollment data until three months after the end of an academic year, job-related data cannot be processed and displayed until at least 18 months after the student exited postsecondary education.

Preparing for the Sub-Committee Meeting

This paper raises a number of questions regarding how best to proceed in structuring information for phase one of the California data system. At the February 2020 meeting, subcommittee members will be asked to consider which data structure would be best for a P20W data set with associated dashboards and query tools, as well as a data structure that would support a process for sharing additional data points upon request. At the March 2020 meeting, once the Workgroup has identified tools that could be developed for individuals and practitioners, the subcommittee will expand its analysis to address the third strategy.

The February meeting will focus on the following questions:

- What are the trade-offs of centralized and federated data structures when considering the strategies outlined for phase one of the California data system,

particularly regarding the effort to align data, view historical data, ensure security, and manage governance complexity?

- How would these structures influence factors such as the effort required to align data, the ability to view historical data, recency of information, security controls, and the complexity of the governance structure?
- Will the same data structure work for both a P20W data set and a process for requesting additional information?
- What topics should be covered in a professional development opportunity that will be held March 9, 2020 and include representatives from other states?

References

Institute of Education Sciences. (2019). *Issue brief: Effective uses of data to gain support for an SLDS*. Washington, DC: Institute of Education Sciences. Retrieved from <https://slds.grads360.org/services/PDCService.svc/GetPDCDocumentFile?fileId=35153>

Levantoff, J. (n.d.). *Costs of state longitudinal data systems*. Washington, DC: Workforce Data Quality Campaign. Retrieved from https://m.nationalskillscoalition.org/resources/publications/file/Cost-of-State-Longitudinal-Data-Systems_web.pdf

Moore, C., & Bracco, K. (2018). *A hunger for information: California's options to meet its statewide education data needs*. Sacramento, CA: Education Insights Center. Retrieved from <http://edinsightscenter.org/Portals/0/ReportPDFs/hunger-for-information-final.pdf>