

Joint Policy & Analytics and Practice & Operations Advisory Groups Meeting Summary

March 11, 2021

The Policy & Analytics Advisory Group provides a means for the public to offer recommendations to the Workgroup about how to ensure the California Cradle-to-Career Data System supports research, evaluation, accountability, and optimization of publicly funded services at the state level.

The Practice & Operations Advisory Group provides a means for the public to offer recommendations to the California Cradle-to-Career Workgroup about how the data system could address improvement efforts at the institutional and regional level, support a case management approach to service delivery, and create tools that would be useful to students, families, and teachers.

This document provides a summary of the key points that emerged from a half-day meeting that provided an update on workgroup decisions, gathered input on draft recommendations for data points and deidentification, shared personas and messaging, brainstormed key competencies for the executive director job description, and discussed the planned content for the April 2021 Legislative Update.

More information about the meeting, including a recording, materials referenced during the meeting, and the PowerPoint, are available at <https://cadatasystem.wested.org/meeting-information/policyanalytics-advisory-group>.

The following advisory group representatives attended the meeting:

Policy & Analytics

Kristin Schumacher, CA Budget and Policy Center; Su Jin Jez, California Competes; Evan White, California Policy Lab; Brian Guerrero, California Teacher's Association; Samantha Tran, Children Now; Emily Putnam-Hornstein, Children's Data Network; Christopher Nellum, The Education Trust-West; Orville Jackson, GreatSchools.org; Heather Hough, PACE; Liz Guillen, Public Advocates; Jacob Jackson, PPIC; and Angela Perry, TICAS

Practice & Operations

Craig Hayward, Bakersfield College; Jessie Ryan (for Sara Arce), Campaign for College Opportunity; Susan Savage, Child Care Resource Center; Anthony Dalton, Futuro Health; Roneeta Guha, Linked Learning Alliance; Kathleen Porter, Poway Unified School District; Nabil Shahin, Santa Clara County Office of Education; Lange Luntao, Stockton Unified School District; Heddy Nam (for David Rattray), UNITE LA; and Tyler Wu, uAspire

Update on Key Decisions by the Cradle-to-Career Workgroup

The meeting opened with Kathy Booth of WestEd providing an update on legislation pertaining to the Cradle-to-Career Data System, including the inclusion of \$18.8 million in the Governor's January budget and trailer bill language that aligns with recommendations from the December 2020 Legislative Report. In addition, Assembly Member Irwin introduced AB99, which also largely reflects workgroup recommendations. The Senate held a hearing on the Cradle-to-Career Data System on March 1 and the Assembly will held a hearing on March 16.

Then, Kathy Booth of WestEd described decisions made by the workgroup at its February meeting, including approving the definition of personally identifiable information (PII), the data classification protocol, the system disclaimer, and the privacy policy.

In addition, she answered a question posed by Anna Alvarado of the EDGE Coalition about whether third parties can contribute their information to the data system. Kathy Booth of WestEd clarified that external records can be matched as part of a study or evaluation using individual-level data, as outlined in the comprehensive data request process. However, this information would not be included in the dashboard or query builder.

The only third-party data source currently flagged for inclusion is National Student Clearinghouse information related to enrolling in college outside of California. However, other data sources could be included in future phases. For example, the workgroup explored the idea of linking the Cradle-to-Career Data System with regional data warehouses but thought this concept should wait for a later stage of development.

[Personally Identifiable Information and Deidentification Protocol](#)

Evan White of the California Policy Lab expressed concern that characterizing personal characteristics as PII will mean that the data providers could decline to release data on key factors such as a student's race, thus undermining the intent of the data system to address equity gaps.

Kathy Booth of WestEd clarified that the PII definition flags elements that must be treated with additional care, rather than indicating that they cannot be shared. The definition will be used in tandem with the deidentification protocol, which describes the way that sensitive information will be protected in public facing tools.

Evan White of the California Policy Lab noted that he was supportive of the deidentification policy, which mirrors the way the California Health and Human Services Agency (CHHS) protects sensitive information. However, he felt that personal characteristics should be removed from the PII definition. Specifically, he noted that while the Health Insurance Portability and Accountability Act (HIPAA) defines any characteristic that could uniquely identify the individual as sensitive, it does not specifically list race, gender, or other similar items. He wants to make sure that the lowest common denominator of HIPAA, the Family Educational Rights and Privacy Act (FERPA), and the Information Practices Act (IPA) are used, and not anything more expansive.

Heather Hough of PACE agreed that if the threshold is anything that can be used to identify an individual then no information can be included. She was uncomfortable with references to "any other unique characteristics" and "unique identifiers" in the list.

Emily Putnam-Hornstein of Children's Data Network argued that the PII definition recommended by the workgroup is not aligned with HIPAA, FERPA, or the IPA. She was concerned that this definition would bring the development of the data system to a halt.

Susan Savage of CCRC agreed and expressed concern that the Legal Subcommittee was undermining agreements made by their colleagues as part of the planning process. Samantha Tran of Children Now concurred.

Orville Jackson of GreatSchools.org was also worried by the expansive definition, particularly because he has experienced difficulties in securing data due to legal constraints. He noted that given that attorneys are tasked with risk mitigation, they are inclined to only share items that are mandated.

Kathy Booth of WestEd noted that the PII definition had been developed by the Legal Subcommittee, which is made up of HIPAA, FERPA, and IPA experts. They established the definition to ensure that their agencies can legally participate in the data system, with the understanding that the deidentification policy would identify how the information can be shared. She indicated that the trailer bill refers to the elements in the P20W data set that shall be provided by the partner entities.

Craig Hayward of Bakersfield College received confirmation from Kathy Booth of WestEd that the deidentification policy would be used to create suppression rules so that the data points could appear in the query tool. He noted that it will be important to push back against the lack of vision that comes from risk managers, as they can undermine the utility of the system. He wondered if there were cases where individuals had been identified after their data was placed in a longitudinal data system, and if this had caused them harm, or if this was a hypothetical concern.

Anthony Dalton of Futuro Health shared the concern about the PII definition, noting that while it will not stop the project from moving forward, but it will severely limit the problems that the public will be able to solve with the data. Su Jin Jez of California Competes, Roneeta Guha of Linked Learning Alliance, Kathleen Porter of Poway Unified School District, Jacob Jackson of PPIC, and Jessie Ryan of the Campaign for College Opportunity agreed.

[Post meeting note: concerns about the PII definition were shared with the Legal Subcommittee. They clarified that while different terms, such as “personal information,” “personally identifiable information,” and “sensitive information” may be used by HIPAA, FERPA, and the IPA, the impact of flagging these data points as protected is the same. Namely, these data points must be subject to the deidentification protocol to ensure that individual privacy is protected and that the partner entities do not violate federal and state law when participating in the data system. Legal Subcommittee members also noted that the full PII definition clarifies that these specific data points are listed because they may cause an individual’s identity to be revealed, not necessarily in isolation, but when combined with other elements. Finally, they affirmed that the data classification scheme reflects the requirements of the IPA.]

Kathy Booth of WestEd described the proposed deidentification protocol.

Heather Hough of PACE asked where the policy would be applied, and Kathy Booth of WestEd clarified that it would be used in the dashboard, query builder, and expedited data request process. It would not be applied to individual-level anonymized data made available through the comprehensive request process. For research conducted using individual-level data, results would be examined by experts through a disclosure review process that is based on rules that apply to the specific data points that were included in the study—this approach allows the Cradle-to-Career Data System to comply with the deidentification policy created by CHHS.

Heather Hough of PACE expressed concern that the suppression rules could create confusion, particularly if people accessed information at both the school site and district level and information available at the district level cannot be displayed at the school level due to small n sizes. It will be important to clarify why information might be available at one level but not at another.

Craig Hayward of Bakersfield College was supportive of the rule about expressing figures as whole numbers.

Wesley Whitaker of Assembly Member Irwin's Office noted that it is important to describe reasons for suppression, so the public understands that the Cradle-to-Career Data System is not hiding important information.

P20W Data Points

Kathy Booth of WestEd described the work done by data experts from the partner entities to refine the list of P20W data points and to assess the quality of the underlying data to inform their inclusion in the data system. This has resulted in a significant increase in the number of data points that will be freely available in the dashboards and query tool and through the request process, such as K-12 science assessments, eligibility for social services in addition to identifying who had received social services, and additional details on English language learners in the community colleges.

She then noted the data points that had been removed from the list because data are not available or that would be available by request only.

a-g Courses

Evan White of California Policy Lab asked for clarification on the removal of the number of a-g courses that students took, given that this data point is in CALPADS.

[Post meeting note: in a subsequent exchange with CDE, they clarified that this data point is not validated. Given analysis that shows significant misalignment between local data systems and the a-g eligibility database maintained by UC, this data point is not of sufficient quality to include in the Cradle-to-Career Data System. However, with the upgrades to CALPADS planned as part of scaling CalifornaiColleges.edu, CDE anticipates that data quality will improve and the data point can be included in the future.]

Early Learning and Care Part-Time/Full-Time Enrollment

Susan Savage of CCRC asked for clarification on the removal of the institutional characteristic regarding whether early learning and care programs enroll students part-time versus full-time, given that she understood this data point to be available at the state level.

[Post meeting note: in a subsequent exchange with CDE, they clarified that it is not possible to reliably link part-time and full-time enrollment to specific providers, but that part time/full time enrollment in early learning and care will be available at the student level.]

English Language Proficiency Assessment

Samantha Tran of Children Now asked why the K12 English Language Proficiency Assessment is listed as available by request only when it currently on the California School Dashboard. Cindy Kazanis of CDE clarified that that item had been mislabeled and would be available on the query builder.

Course-Level Information

Craig Hayward of Bakersfield College asked whether the K-12 Course Group State Codes and course titles will be included. Kathy Booth of WestEd indicated that course codes would be available for career and technical education courses in the query builder tool.

[Post meeting note: course titles for middle school math courses will be available through the request process.]

Craig Hayward encouraged the data system to include comprehensive course-level data in future years. For example, it will be helpful to examine the specific high school math courses that students take and how that prepares students for success in college math curriculum.

Location of Employer Headquarters

Kathy Booth of WestEd explained that the location of an employer's headquarters had been removed because it is not an accurate proxy for where students are employed. Evan White of the California Policy Lab argued that, given that this data is of high quality and may be useful for analyses about employment outcomes, the data providers should not constrain access to data just because it might be confusing for some users. He suggested renaming the data point to reduce confusion. Su Jin Jez of California Competes and Jacob Jackson of PPIC agreed.

Craig Hayward of Bakersfield College concurred. He noted that data are generally imperfect and researchers spend a lot of time cleaning and improving the data set for analysis. Typically, incomplete data or missing data can be excluded from the analysis and the remaining data are still very valuable. He suggested that the employer headquarters be made available by request. Su Jin Jez of California Competes underscored the value of assigning complex data points to the "request only" category rather than removing them.

[Post meeting note: in a subsequent exchange with EDD, they are willing to consider making the data point available by request. They will gather additional information about the nature of the data point and how it is used by the Labor Market Information Division to inform this decision.]

Grade Point Average (GPA)

Kathy Booth of WestEd explained that the data providers had removed this data point because it is not currently posted on public-facing tools by the partner entities. Many had concerns about the quality of the underlying grade data and felt that it did not meet the threshold for inclusion in the Cradle-to-Career Data System. However, they noted that grade data could be prioritized for cleanup so that they can be used to calculate GPA in the future.

Nabil Shahin of Santa Clara County Office of Education agreed that GPA is an imperfect measure because there is a high level of subjectivity. However, he was concerned that not including GPA would mean that high school information would be overly focused on assessments. Evan White of California Policy Lab and Su Jin Jez of California Competes shared this concern.

[Post meeting note: in addition to assessment data, the following data points on high school students will be included in the query builder: ninth grade math course, number of math courses completed in high school, highest K-12 math course completed, completed an AVID course, number of Advanced Placement courses completed, number of International Baccalaureate courses completed, number of career and technical education courses completed, completed a career and technical education pathway, subject matter of career and technical education pathway, completed a work based learning program, completed a distance learning course, met a-g course eligibility requirements, completed a college course while in high school, completed a leadership/military science course, earned the State Seal of Biliteracy, individual College and Career Indicator level, and type of high school completion.]

Craig Hayward of Bakersfield College noted that high school GPA was the most essential and critical data element in transforming the California Community College system's approach to assessment and placement in math, English, and ESL. Evan White of California Policy Lab, Su Jin Jez of California Competes, Jacob Jackson of PPIC, Jessie Ryan of Campaign for College Opportunity, Orville Jackson of GreatSchools.org, and Wesley Whitaker of Assembly Member Irwin's Office agreed.

Craig Hayward of Bakersfield College, Nabil Shahin of Santa Clara County Office of Education, and Emily Putnam-Hornstein of Children's Data Network requested that annual cumulative GPA be included in the data system.

Cindy Kazanis of CDE clarified that her agency does not currently calculate GPA but does collect grades. Her organization is willing to work to improve the underlying data quality so that GPA can be included in the future. She noted that the proposed updating of CALPADS to integrate with CaliforniaColleges.edu will make this possible.

Evan White of California Policy Lab argued that if there are concerns about underlying data quality, this could be addressed when results of studies are released, rather than limiting access to information.

Orville Jackson of GreatSchools.org and Tyler Wu of uAspire noted that CSAC has GPA information. Evan White of California Policy Lab shared that CSAC gets a list of students with GPAs >3.0 and >2.0, during students' senior year.

[Post meeting note: in a subsequent exchange with CSAC, they clarified that the GPA they receive from local education agencies (LEA) is not a comprehensive GPA—some LEAs do not report to CSAC, specific types of courses such as physical education are excluded, and information is generally only available through 11th grade. As part of Cal Grant modernization, CSAC has proposed eliminating GPA from state aid eligibility requirements. Furthermore, the trailer bill specifies that by 2025, GPAs used by CSAC should come from CDE rather than from LEAs. However, CSAC did agree to reconsider having duration of financial aid be an item available by request only and determined that a data point should be included in the query builder that shows the proportion of terms in which a student was enrolled in postsecondary that they received financial aid.]

Kathleen Porter of Poway Unified School District cautioned that GPA is not straightforward. For example, individual institutions weight GPA differently, both in the value reported by K-12 institutions (some have a weighted GPA for more demanding courses) and in how postsecondary institutions weigh specific courses when examining GPAs for the purpose of admitting students. Also, the GPAs that are calculated for various purposes will not align with transcript-level information.

Craig Hayward of Bakersfield College noted that researchers do not need transcript-quality GPAs for studies. The Cradle-to-Career Data System could include an "estimated GPA" for research purposes. Evan White of California Policy Lab and Roneeta Guha of Linked Learning Alliance agreed.

Financial Aid

Kathy Booth of WestEd noted that a number of the postsecondary partners are concerned that they cannot share financial aid information that originates from the FAFSA due to guidance released by the U.S. Department of Education's Privacy Technical Assistance Center in 2017. WestEd is seeking clarification from the federal government on allowable use. The advisory groups had no questions or comments.

Opt Out Language

Kathy Booth of WestEd described the opt out policy and the language that would be used to alert people to this option. The advisory groups had no questions or comments.

Personas and Key Messages

Kathy Booth of WestEd described the work of the Community Engagement Subcommittee to document the types of people who will be likely users of the data system, which are codified in 23 personas. Each persona has been given a name, picture, and a demographic profile and includes the types of challenges they face, what they are looking for, and what they need to be successful. The personas are grouped into four archetypes—analyzers, planners, practitioners, individuals.

Liz Guillen of Public Advocates noted that specific people can sometime fill multiple roles. She was glad that there is a separate persona for individuals that might access the data system, and noted that it will be vital to craft appropriate communications for them.

Brian Guerrero of the California Teacher’s Association stated that he was glad that the teacher/practitioner role was recognized. Samantha Tran of Children Now agreed. Susan Savage of CCRC received clarification that early care providers would also fit into this category.

WestEd conducted a poll to determine which of the four archetypes the advisory groups resonated with the most. The majority identified as analyzers. Kathy Booth of WestEd noted that it is important to recognize the perspectives that have been most strongly represented in the planning process and those who have had a smaller role. This imbalance can be addressed in the user centered design process.

Kathy Booth of WestEd then described the audience matrix developed by the Community Engagement Subcommittee. Organized by person type, the document lists goals that would be supported by information from the data system, which tools they are most likely to use, core messages about those tools, and how best to communicate with them. This information will be used to craft a communications plan. Finally, she noted that the personas and the audience matrix are being vetted more broadly with communities that embody the archetypes.

Liz Guillen of Public Advocates shared that the audience matrix is in a format is not easy to share with the people that fit within the persona categories. She was concerned that people might not see themselves in the categories as described.

Lange Luntao of Stockton Unified School District noted that advocates seemed to be missing from the list, which may exist at the nexus of analyzers and planners. He agreed to work with WestEd to develop a persona that fits this profile.

Executive Director Job Description

After Kathy Booth of WestEd clarified that the executive director will be hired by the governing board, and so does not need to fit into standard state classifications, the group used a Google document to brainstorm key characteristics for the executive director, which can be used to craft a job description. Key attributes included:

- Communications experience
 - Government relations
 - Public relations

- Diplomacy
 - Someone who many parties are comfortable working with and has credibility with advocates
 - Able to balance interests of data providers and data requestors
 - Able to lead in complex multi-stakeholder environments, particularly with the governing board
 - Able to lead by influence, given that the role doesn't hold authority/power to mandate change but will need to weigh the opinions/positions to figure out a way forward
- Technical expertise
 - Analytics/research
 - Data visualization
 - Education data
 - Data policy
 - Experience as an educator in or graduate of a California public education entity
 - State/government processes
 - Race and gender justice
 - Serving diverse populations
 - Providing professional development and technical assistance (this is a nice-to-have)
 - Legal (this is a nice-to-have)
- Vision
 - Able to innovate and move the vision forward
 - Has an education equity theory of change
 - Willing to name problems that the data reveal
- Management
 - Can build a team from the ground up
 - Can leading diverse teams

The group noted that this is a long list of qualifications. The group wanted to ensure that women and people of color would apply.

When asked to prioritize items from the list, these two items were listed the most:

- Demonstrated commitment to data as a tool for advancing educational equity
- Balancing the needs of multiple, diverse stakeholders

April Legislative Update

Kathy Booth of WestEd walked through the planned outline for the report. It will be shared with the workgroup after the March 25 meeting, as much of the content will hinge on decisions made at the meeting. It is due to the legislature on April 1.