Policy & Analytics Advisory Group Meeting Summary

April 14, 2020

The Policy & Analytics Advisory Group includes a broad range of perspectives and 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.

This document provides a summary of the key points that emerged from substantive discussion over the course the day. 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:

Liza Chu, Asian Americans Advancing Justice; Heather Hough, Policy Analysis for California Education; Jacob Jackson, Public Policy Institute of California; Su Jin Jez, California Competes; Carlise King, Child Trends; Rigel Massaro, Public Advocates; Brian Guerrero, California Teachers Association; Angela Perry, The Institute for College Access and Success; Emily Putnam-Hornstein, Children's Data Network; Christopher Nellum, The Education Trust-West; Kristin Schumacher, California Budget & Policy Center; Samantha Tran, Children Now; Andrea Venezia, Education Insights Center; and Evan White, California Policy Lab.

Update and Agenda Review

The meeting opened with a high level summary of the three use cases under discussion, how the use cases were developed, and a clarification regarding the differences between the information that would potentially be contained within the P20W Data Set versus accessed through the Data Request process.

(Note: Ben Chida from the Governor's Office was not able to join the group until after lunch, but his update on the process is included here).

Ben Chida assured that group that the pandemic has redoubled the Governor's Office commitment to designing a state data system. Access to data is essential when managing a crisis, and a stronger intersegmental system will make the state more resilient and adaptable. Reflecting on the current situation, the data system will be needed to understand how policy changes impacted equity for those who are most in need and least able to bear the impact of pandemic. In the future, it would be helpful to better understand what the impact of policy changes might be before they are made.

When asked how much the advisory group should weigh future budget constraints in deliberations, Ben Chida responded that the group should backwards design from the system they would like to have in the future. As part of this process, they should identify which components would provide a strong foothold in responding to state needs in the short term. It is important to provide the full vision and to clarify which components are most urgent to create first.

When responding to a question on the challenge of engaging likely data users about the design in the midst of the pandemic, Ben Chida recommended that the group leverage their expertise to keep up momentum, rather than wait until all stakeholders can be fully engaged.

Providing a national perspective, Paige Kowalski from the Data Quality Campaign noted that states with longitudinal data systems are currently scrambling to put these systems to work as they respond the crisis and are seeking to invest more in their data—a response that is similar to what happened during the Great Recession (2007 to 2009). She also noted that other states are starting to contemplate the types of tools for practitioners and individuals that are part of the California use cases.

P20W Data Set Use Case

After reviewing key components of the P20W data set use case, the group provided input regarding its scope and focus. Several advisory group members commented on the specific data elements. Some flagged the need to provide more nuance related to suggested variables, particularly regarding disability and language (which should distinguish between home language, spoken language, English language learner, and dual language learner). Parental education, special education status, and attendance/chronic absenteeism were also listed as desired variables for the P20W data set. Other topics were identified as important, but potentially too sensitive to include in the P20W data set, such as immigration status.

One key area of conversation was the types of early care data that should be included, especially given a parallel effort to build an early childhood integrated data system (ECIDS) and the fact that early care information is held by both the California Department of Education (CDE) and Health and Human Services (HHS). One participant felt that linking CDE's early care data to K-12 information would be a good first step, if additional data sets could be added at a later date. Another participant suggested that the state data system leverage information managed by DSS that could provide a yes/no flag regarding whether an individual had received childcare assistance, as a way to expand information in phase one. A third participant reflected on the way Minnesota uses the same match engine for both its ECIDS and P20W system, but produces separate dashboards. This approach might work well for California, but it would require that the ECIDS stakeholders be consulted.

Another area of discussion was employment data. Several advisory group members encouraged the state to begin work on expanding available information as soon as possible, so that additional data points or sources could be included in phase two. Having more granular employment and earnings data will become even more important as the state seeks to understand the impact of the pandemic. Specifically, members flagged the value of information from the Franchise Tax Board related to self-employment, student loans, and linkages between parents and children. The group also discussed the value of having employers report additional data points for the Unemployment Insurance file including Standard Occupational Codes and hours worked per quarter.

Finally, the group discussed the proposed financial aid data. Given the value of understanding student debt, one participant suggested tracking the cost of college programs. However, the state will need to determine the best methodology, such as net price calculators or gainful employment definitions. One participant noted that it would be valuable to get a comprehensive list of financial aid received by students. In addition to collecting some of this information from colleges, the state could invest in bulk-purchasing information, which is available from credit bureaus, and could be used to understand how college debt impacts other factors like home ownership and family formation. Having the state negotiate this purchase could provide an overall cost savings, much in the same way the state could invest in a single purchase of information on out-of-state college enrollments from the National Student Clearinghouse.

Tools to Access the P20W Data Set

First, the group was introduced to three user stories, which use the characteristic of comfort with quantitative data to create the categories of data novice, data apprentice, and data expert. Then, advisory group members broke into small groups to discuss three types of tools that could be used to access the P20W data set: dashboards, query tools, and firewalled analytical tools.

After extensive discussion (highlights below), the group developed the following vision:

P20W data should be available to the public through two types of tools:

1) Dashboards that present simple, user-friendly information, with the ability to drill down to more sophisticated views and features, such options to compare results, disaggregate by student categories, and export files containing summary data
2) Query tools that allow users to build a combine variables from a list of elements and export results as charts or summary data

In addition, a secure data environment should provide authorized users with access to de-identified, unitary data. The information available through this environment would include both the P20W data set and more extensive data from all partner entities. This data set should be used in three ways:

1) Conduct approved research studies
2) Allow external partners to implement work on behalf of the partner entities
3) Construct data sets necessary to produce additional tools such as dashboards

Among these options, the scale could be adjusted based on available resources. Regardless of the scope of phase one, the technical and governance underpinnings for the state data system should allow for all of these options to be developed over time.

Discussion on Dashboards

The small group recommended that dashboards follow a simple to complex design, intended for the apprentice-level user and offered in several languages. The front end should provide information in a clear, intuitive, and engaging manner, and allow users to look deeper such as showing results by region or over time. The dashboard should be accompanied by a strong communications plan that helps users understand its goals and gathers input on how to grow features over time. In a reflection of this desire to create a dashboard that offers both simple and complex features, votes on the desired dashboard style were split, with 50% (7 people) supporting an infographic-based approach, 29% supporting simple charts (4 people), and 14% supporting complex charts (2 people), and one person abstaining.

Discussion on Query Tools

The small group recommended that the state provide a query builder that follows a complex to simple design, intended for the apprentice-level user. The query tool could provide different information than what is in the dashboard. Users should be able to determine which variables to combine from an easy to understand list (such as elements, timeframes, geographic regions, and subpopulations) and have the ability to export both summary data and charts. The tool could also include a simpler interface that allows users to fill in pre-written sentences or provide a set of pre-created tables for common topics. The query builder tool could provide a cost savings to partner entities because members of the public

could generate their own answers for simple data requests. Ten members (71%) voted for a query builder approach, two (14%) selected web-based tables, and two abstained.

Discussion on Firewalled Analytical Tools

The small group recommended that this tool focus on data experts, with the state providing authorized users with access to unitary data. These data could be used to build tools like dashboards in addition to conducting analyses. Providing access to unitary data is an important complement to dashboards and query builders because it is difficult to conduct in-depth analyses when working with pre-established aggregate data sets. Many of the questions that policy makers and the public are likely to ask—such as any question related to causality—will require analyses based on unitary data, particularly when looking outside of accountability frameworks. Having unitary data behind a firewall is important to ensure that information can be provided on all populations without the risk of their identity being exposed. It is also important to have a process to review what people intend to do with (and how they interpret) information, to ensure that the information coming from the data system is of high quality. HHS has already built this type of tool, which could be expanded to include education, financial aid, and employment data. The 13 members who voted supported access to unitary data (one person abstained).

Data Request Use Case

After reviewing the concepts within the data request use case and how they relate to the proposed tools, the group considered the guideline that only linked data could be accessed through the state data system—meaning that information would only be provided for analyses that combine information from more than one partner entity. Some participants raised concerns about this requirement. Some felt that the state data system should provide access to both linked and unlinked data. One noted that in the absence of clear and consistent data request processes across partner entities, researchers might be incentivized to ask for linked data as a way to get access to information from within a single agency. The group consulted with Paige Kowalski from the Data Quality Campaign, who clarified that other states only share linked data and would consider the alternative a deal-breaker. Some advisory group members suggested that the partner entities consider ways to streamline or align request procedures. Another noted that the partner entities could elect to use the state system data to manage their request process as a way to reduce their workload.

Tools for Practitioners and Individuals

The discussion began with a snap poll of both advisory group members and members of the public who were attending the Zoom meeting about whether tools for practitioners and individuals should be included in phase one. About three-quarters of respondents reported that they should. When those who voted no were asked why, reasons included:

- it will be difficult and costly to get data that is sufficiently recent and accurate to populate these tools
- the state should focus first on created the linked data set before creating tools, to determine which information can be reliably integrated into public-facing tools
- scaling up existing tools is not a strategic approach for the state data system

Then the group reviewed the features of the three possible tools, and voted on which ones should be included in phase one.

CaliforniaColleges.edu, hosted by the California Colleges Guidance Initiative (CCGI)

Nine members of the group (62%) voted to recommend this tool, with three no votes (21%) and two abstentions. One member noted that this option qualifies as low-hanging fruit because it is an existing system that could quickly generate useful student-level information. Another member indicated that providing valuable information to those working most closely with students would build confidence in the state data system. Those who voted no felt that this type of tool, while valuable, was not a priority for the state effort in the first phase and would be too costly.

eTranscript California, hosted by the California Community Colleges Chancellor's Office, in partnership with the California Department of Social Services (CDSS)

Seven members of the group (57%) voted against recommending this tool, with one yes vote and five abstentions. Concerns echoed the initial response about the wisdom of investing in tools for practitioners and individuals. One person felt that eTranscript California was less useful than CaliforniaColleges.edu because it lacks planning and degree audit tools.

Secure Data Environment, hosted by the Silicon Valley Regional Data Trust, in partnership with Stanislaus Cradle-to-Career Initiative

About forty percent recommended this service (with one "yes" and five "yes, with reservations") and eight (57%) voted no.

One advisory group member indicated that more information was needed about the state's role in the project. Another noted that there are significant privacy concerns among parents in Silicon Valley about how the system allows access to sensitive information. This concern was echoed by another participant who felt that passing sensitive student-level information between agencies would be seen as too risky in the first phase of this project. Finally, a participant felt that the underlying technology was outdated and should not be scaled.