

Definitions Subcommittee Meeting Summary

June 9, 2020

This document provides a summary of key points that emerged over the course of a day-long meeting. More information about the meeting, including the background paper, data element documentation template, PowerPoint, and a meeting recording are available at <https://cadatasystem.wested.org/meeting-information/definitions-subcommittee>.

The Definitions Subcommittee will document technical definitions for key information that will be shared between partner entities in Phase One. The June 2020 meeting had the following goals:

- Clarify the purpose of the California data system and the recommended scope for phase one
- Determine the process we will use to align data definitions for public display
- Identify key issues for aligning definitions for:
 - Race/ethnicity
 - Sex/gender
 - Sexual orientation

The following representatives attended the meeting:

Randy Tarnowski, Association of Independent California Colleges and Universities; Joanna Murray, Bureau for Private Postsecondary Education; Todd Hoig, California Community College Chancellor's Office; Jerry Winkler & Channa Hewawickrama, California Department of Education; Janet Buehler, California Department of Technology; David Sanabria & Jennifer Schwartz, California Health and Human Services Agency; Patrick Delaney, California Department of Social Services; Monica Malhotra, California State University; Rima Mendez & Amy Tong, California School Information Services; Adrian Felix, California Student Aid Commission; Erin Skubal, California Teacher Credentialing Commission; Margo Gonzalez, Employment Development Department; Patrick Getz & Daniel Rounds, Labor and Workforce Development Agency; Chris Furgiele, University of California Office of the President; Todd Britton, University of LaVerne; and Valerie Mendelsohn, West Coast University

Workgroup Update

The meeting opened with the facilitator outlining the planning process and providing a summary of decisions made by the Cradle-to-Career Workgroup to date.

Process for Aligning Data Definitions

After reviewing the background paper that described strategies for aligning data definitions and showed disparities in how the education segments in California define and display the concepts of race/ethnicity, sex/gender, and sexual preference, the subcommittee highlighted key items to consider for aligning definitions in public displays of Cradle-to-Career Data System information. Ideas that emerged from the discussion included:

- Some differences may be about labeling, but not definitions, so the process should distinguish between these areas of divergence

- Many individual institutions and some partner entities collect information that is not displayed publicly—for example, some data elements may be available in specific years but may not have attained maturity sufficient for public display
- Because demographic information is self-reported and voluntary, designations will be missing for some students, so the process needs to distinguish between individuals who declined to state information versus instances where the data are not present
- To build trust in the data, it may be helpful to clarify the origin of each item and describe how information gets adjusted as it moves from local institutions to state agencies to the Cradle-to-Career Data System
- The user interface for the dashboard and query tool should be constructed in a way that helps users understand the complexities of the data without deterring them from using it
- Representatives from several partner entities indicated that they only collect categories required for Integrated Postsecondary Education Data System (IPEDS)—particularly independent colleges that do not have a centralized statewide system—but some noted that it would be helpful to have a list of other elements to collect that is normed across education providers and other state agencies

The group discussed the importance of beginning with a set of policies and principles for aligning definitions before tackling specific categories of information. Some initial recommendations included:

- Articulate the intent for aligning information
- Provide information at the most granular level possible, while ensuring that data suppression keeps information on individuals secure
- Identify opportunities for expanding categories in future phases, but begin with elements that most partner entities collect
- Wherever possible, adopt California definitions over federal definitions, and clarify why different definitions are being used
- Create a consistent process for handling missing information
- In cases where partner entity data are being grouped into a broader category, create a translation table, clarify when one status trumps another, and indicate when specific categories are being assigned to an “Other” category
- Acknowledge when specific partners are not collecting data in the categories displayed in public tools
- In public tools, note when data definitions changed in a manner that is easy to find, such as in the Minnesota dashboard
- Determine how far back to go for displaying information
- Create a process for documenting changes to definitions

Finally, the group noted items that could be improved in the documentation form:

- Document effective and end dates for each element (and iteration of each element)
- List usage rules related to how information could be displayed in public dashboards and query tools

Contextual Information on Race/Ethnicity Data

Note: The information in the following three sections reflect topics discussed in small group break outs, as well as the full group discussion.

Patrick Getz and Dan Rounds from LWDA offered their perspectives on aligning race/ethnicity data for the CAAL-Skills data system, which links records across Workforce Innovation and Opportunity Act (WIOA) training providers. They found that the way race and ethnicity data are collected—such as how many questions are asked and the order in which those questions are asked—has a strong impact on what gets reported. One way to address the differences is to provide results in several ways, such as to include separate reports for race and for ethnicity, and to also show results based on the categories used by each participating agency. They also noted that when data are aggregated at too high a level, such as with the broad categories of “Asian” or “Multi-race,” the experience of specific groups may be obscured.

One participant argued that it would be preferable to use the IPEDS categories. While they are more limited than the information collected by some partners, IPEDS categories are available from all of the education partners and are of stronger quality because they have been collected in a consistent manner for a decade.

Others agreed that the IPEDS categories are a good starting place but wondered if it would be possible to allow data to be drilled down to more specific categories when available. This could help surface available data and make a case for more comprehensive categories. However, another participant noted that any change to definitions must take into account the time and expense needed at the local level to adjust data collection and storage systems.

One person suggested that Cradle-to-Career Data System definitions should be published to California’s Open Data Portal so those definitions can be referenced by vendors, researchers, and the public. Making the definitions available will also encourage discussion and provide a feedback loop.

Finally, another participant noted that if information is displayed before 2010, when the federal definition of ethnicity changed, it will be important to clarify that time trends will be affected by this change.

Contextual Information on Sex/Gender Data

The subcommittee discussed the challenge of the federal definition for sex and gender being more restrictive than what is required by California law. For some institutions, there is an increasing number of students reporting nonbinary gender categories, which is creating challenges given that individuals must be reassigned to allowable categories for federal reporting. Furthermore, guidance provided through entities like IPEDS about how to address this issue are not compatible with how data are handled in longitudinal data systems (such as allowing for free text field responses). CSU and UC shared the different ways in which they reassign nonbinary students to binary sex categories for IPEDS reporting.

One participant noted that non-binary categories are relatively new, which means that data may not yet be mature. Furthermore, the data may be suppressed in some cases due to small cell sizes.

Another noted that if there is a placeholder for nonbinary, even if it is not yet populated, this could help to drive future availability. For example, education institutions would be aware that they need vendors to provide additional gender categories.

One additional possibility would be to add an option to the query builder that allows users to select either California definitions (which would include a nonbinary option) or federal definitions (which would have only male/female). Adopting federal definition will require a logic to reassign nonbinary to binary sex categories, which will need to be coordinated with each partner entity, as methodologies vary by agency.

An informal poll using the chat feature showed that most subcommittee members preferred adopting California's definition (which includes a nonbinary option), even though data would be missing from some institutions and suppressed in some cases to protect individual identities. However, one participant noted that the absence of federal definition or variability in data element values will restrict comparison among the states.

Contextual Information on Sexual Orientation Data

Many education institutions do not capture sexual orientation, and CDE is prohibited from doing so. Because CHHS can only collect data necessary for delivering services, many programs do not collect sexual orientation. One participant questioned the ethics of collecting this information and another noted that legal requirements on collecting sexual orientation are still evolving.

One participant described an effort to add a data element on sexual orientation that was complicated by discrepancies in recommendations from advocacy organizations versus researchers about the appropriate categories to include.

Another noted that having two options—one for heterosexuals and one that encompasses all other categories—might be easiest to implement as a first step for the Cradle-to-Career Data System.

Next Steps

Subcommittee members were asked to do the following:

By June 18:

- 1) Send LeAnn Fong-Batkin (lfongba@wested.org) examples of any principles or policies related to topics such grouping definitions or handling collisions that could be used to establish the framework for our July meeting
- 2) Send Kathy Booth (kbooth@wested.org) edits on the meeting summary

By July 7:

Return the worksheet with definitions for race/ethnicity, sex/gender, sexual orientation, with the following modifications:

- for years available, note the effective and end dates, such as May 5, 2010-December 31, 2018
- list usage rules related to how information could be displayed in public dashboards and query tools (for example: this data should not be displayed because the item is too new)
- provide referential information such as: data changed in 2010 due to a shift in IPEDS requirements

The schedule for topics at future meetings is tentatively:

Meeting	Topic	Data Elements
July 2020	Policies for data alignment	
August 2020	Match elements	Race/ethnicity, sex/gender, sexual orientation, name, address, social security number, date of birth, agency identifiers
September 2020	Student characteristics	Age bracket, social service participation, foster youth, socioeconomic status, homeless, migrant, parental education, disabled, special education, military, expected family contribution for college, financial aid dependency status
October 2020	Financial and college application	Financial aid application status, financial aid approval status, expected family contribution, types of aid received, length of aid received, college application status, college acceptance status, transfer application status, transfer acceptance status
November 2020	Enrollment, grades, academic progress, tests, eligibility	K-12 attendance, K-12 absenteeism, college enrollment, on K-12 GPA, grades, UC and CSU eligibility, Smarter Balanced and NGSS scores in 3 rd , 5 th , 8 th , 11 th grades, AP/IB/ACT/SAT scores, postsecondary satisfactory academic progress, transfer preparation
December 2020	Course characteristics, pathways, majors, transfer readiness	K-12 CTE courses/pathways, work-based learning, math courses/pathways, dual enrollment, online, content area, remedial, gateway, transferrable, education goal, declared major
January 2021	Units, milestones, transfer, completion	K-12 CTE pathway completion, K-12 graduation, postsecondary units earned, fall-to-spring retention, third term retention, satisfactory academic progress, transfer preparation, postsecondary award, postsecondary award subject matter
February 2021	Employment	Employed, earnings, living wage attainment, industry
March 2021	Early care, class size, co-curriculars	Class size bracket, TBD
April 2021	Institutions	Type of early care program, type of kindergarten, total cost of college
May 2021	Revisit deduplication rules, timeframes, and other policies	
June 2021	Review draft documentation	