# Technology & Security Subcommittee Meeting Summary

October 6, 2020

This document provides a summary of key points that emerged over the course of the meeting. More information about the meeting, including the PowerPoint and a meeting recording are available at https://cadatasystem.wested.org/meeting-information/common-identifier-subcommittee.

The October 2020 meeting had the following goals:

- Provide an update on key decisions by the Cradle-to-Career Workgroup
- Gather information related to the system architecture
- Identify components for a service level agreement with the managing entity
- Clarify possible models for data classification schemes

The following representatives attended the meeting:

Formeka Dent, Antelope Valley Union High School District; Helen Norris, Association of Independent California Colleges and Universities; Clarissa Serrato-Chavez (for Jason Piccione) Bureau for Private Postsecondary Education; Andy Manguia, California Commission on Teacher Credentialing; Daryl Lal (for Barney Gomez), California Community College Chancellor's Office; Rodney Okamoto, California Department of Education; Karissa Vidamo, California Department of Social Services; Janet Buehler, California Department of Technology; Lloyd Indig, California Health and Human Services Agency; Amy Fong and Greg Scull, California School Information Services; Subash D'Souza, California State University Chancellor's Office; Gurinder Bains, California Student Aid Commission; Todd Ibbotson, Employment Development Department; Jenni Abbott, Modesto Junior College; Matthew Linzer & Hooman Pejman, University of California Office of the President

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

Kathy Booth of WestEd provided an update of decisions made by the Cradle-to-Career Workgroup in the July, August, and September meetings and described the upcoming community engagement webinars.

One subcommittee member asked for clarification about the timeframe for implementation. Kathy Booth noted that the initial workgroup recommendations will go to the Legislature in December 2020, with the committees continuing to meet through the first half of 2021 to create additional documentation to support implementation. Provided that funding is secured, work would begin in July 2021. Baron Rodriguez of WestEd noted that in other states, it takes about a year to get data into a system and implement master data management, after which additional time is needed for testing and validation. As part of the planning process, another subcommittee is documenting how each agency defines the data points that have been recommended for the P20W data set and recommending how to display this information on the public facing tools. Doing this work in advance can help to speed the implementation timeline.

## System Architecture

Baron Rodriguez of WestEd provided background information on secure data enclaves, the mechanism by which third parties would be able to access unitary data for approved Cradle-to-Career data system research requests.

One member of the group noted that even with the many security features built into secure data enclaves, researchers could still take a picture of the screen with their phone. Kathy Booth of WestEd noted that all individuals who would have access to the secure data enclave would sign a legal agreement (such as a nondisclosure agreement) regarding use of the data and would have had to complete training on data usage, such as a CITI certification.

Another subcommittee member noted the importance of implementing multi-factor authentication, to ensure that the appropriate individuals are accessing the data, not just the designated computers. State and federal regulations require multi-factor authentication in cases where personally identifiable information is accessed.

A third subcommittee member questioned why a diagram in the PowerPoint presentation showed a secure data enclave built in Microsoft Azure as opposed to using another provider such as Amazon Web Services. Baron Rodriguez clarified that he was describing a specific secure data enclave that WestEd manages and that Microsoft Azure was selected given the easier integration into Active Directory. However, the Cradle-to-Career system could be built using any approved cloud provider.

Subcommittee members broke into small groups to discuss various components of system architecture. The notes below reflect both the discussions and the report out to the full group.

## Extract, Transfer, Load (ETL) Processes

A small group identified critical issues related to ETL, the process by which information would be transferred from the data providers to the cloud repository managed by GovOps. Topics included:

- create a data schema or structure to format data, particularly as it may come in any form
- check the integrity of the data, to ensure it is not manipulated by bad actors
- provide documentation related to both business rules and programming to support staff transitions
- update documentation regularly
- provide log ins with multi-factor authentication
- ensure adequate testing and develop cases around common problems like changes to data and formatting changes (such as dropped or added columns)
- develop a process to notify the managing entity about data changes
- develop a process to notify the data provider when their data set is not compliant
- create business rules for how to derive data points
- provide an ETL tool for data providers that do not have one
- establish timeframes for when data providers are expected to upload information

## Business Intelligence (BI) Tools

A small group identified key criteria when selecting BI tools. BI tools help to make information in the data system available to users, such as through dashboards and query builder tools. Topics included:

- ensure common features such as ad hoc querying and security
- prioritize accessibility, including addressing cognitive accessibility
- select a web-based tool that is available on multiple platforms such as mobile and desktop
- ensure sufficient capacity to maintain performance even under high levels of usage
- design a user-friendly interface that makes information accessible for a variety of audiences

- integrate technical help functions for the managing entity and public-facing help features that allow users to better understand the information produced
- provide displays that show trends, comparisons, and geographic views
- allow users to filter the information, drag and drop desired components, and drill down to more detailed information
- enable users to pull and export simple charts
- use metadata to track changes to fundamental categories, such as shifts in gender options
- provide a built-in feedback feature to capture information on how to improve the interface or flag potential issues with the data
- regularly update and improve the interface
- treat predictive analytic features with caution, such as by clarifying the accuracy of information produced

The small group recommended that the managing entity use an off-the-shelf BI solution and devote resources to support implementing a user-centered, accessible design.

### Secure Data Enclaves

A small group identified key criteria when developing a secure data enclave to control access to sensitive information. Topics included:

- the number of users that the system should support simultaneously needs to be determined
- there should be proof of data destruction after the conclusion of the approved use
- the managing entity should determine the minimum level of training required, particularly if the Cradle-to-Career data system only provides access to specific analytical tools within the secure data enclave
- the system should use multi-factor authentication
- there needs to be appropriate system and data integrity controls, such as those listed at **https://devblogs.microsoft.com/azuregov/cmmc-with-microsoft-azure-system-information-integrity-10-of-10/**
- the managing entity should implement appropriate deidentification guidelines based on the approved use
- users should sign a rules of behavior document each time they use the system, such as **https://www.wrc.noaa.gov/forms/SEA11003F-OCIO%20Seattle%20IT%20Rules%20Of%20Behavior%20v1_3.pdf**
- there should be a disclosure review board to ensure that no sensitive data are released
- the managing entity should alert data providers if there are changes made to the data or data structures
- wherever possible, the managing entity should provide files with properly deidentified data, rather than providing access to unitary data in the secure data enclave, and there should be clear processes for when access to unitary data is appropriate
- explore the possibility of data licensing for the legal framework

The group identified several secure data enclaves that could serve as a model for the Cradle-to-Career data system including:

- San Diego Supercomputer Center's Sherlock Division

- University of Chicago
- University of Texas

## Service Level Agreement (SLA)

Baron Rodriguez of WestEd reviewed a list of potential topics that could be covered in an SLA for the managing entity, which describes expectations for service delivery on topics such as disaster recovery, documentation, and response times. He identified some items from that list that will be covered in the legal agreement with GovOps. Subcommittee members volunteered to join a homework team to develop an SLA that would be appropriate for the Cradle-to-Career data system, which will be discussed at the November meeting.

## Data Classification Scheme

Baron Rodriguez of WestEd summarized the existing California Department of Technology data classification scheme. Data classification schemes allow specific data points to be flagged based on their level of their sensitivity, such as confidential information that cannot be disclosed in a California Public Records Act request or personal information such as name or social security number. He remarked that, unlike many states, the California data classification scheme does not provide explicit classification for education records. Janet Buehler of CDT agreed to research whether these specifics are provided through other documentation.

Baron Rodriguez described data classification schemes used in several other states. He noted that, in addition to data providers tagging their data, the managing entity will need to classify data once it is combined in the data system because linkages may make information more identifiable—particularly when education and health records are combined.

One participant noted that the entity requesting the data can also influence the level of sensitivity—for example a data point may be flagged as sensitive for a third party but not for a data provider.

A number of subcommittee members shared their data classification schemes, to clarify how education records are handled. One noted that it would be helpful to have a standard format for all education institutions.

Subcommittee members volunteered to join a homework team to determine how the current CDT classification might need to be modified for the Cradle-to-Career data system.