Technology & Security Subcommittee Meeting Summary

January 19, 2021

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

The January 2021 meeting had the following goals:

- Discuss documents created by the Legal Subcommittee including:
  - Data Classification Scheme
  - Privacy Policy
  - Personally Identifiable Information (PII) Definition
  - Deidentification Definitions
- Discuss recommendations from homework teams including:
  - Deidentification Protocol
  - Permissions Protocol

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, Bureau for Private Postsecondary Education; Andy Manguia, California Commission on Teacher Credentialing; Barney Gomez and Daryl Lal, California Community College Chancellor's Office; Alan Nakahara and Rodney Okamoto, California Department of Education; Karissa Vidamo, California Department of Social Services; Janet Buehler and Michele Robinson, California Department of Technology; Lloyd Indig, California Health and Human Services Agency; Greg Scull, California School Information Services; Subash D'Souza, California State University Chancellor's Office; Gurinder Bains, California Student Aid Commission; Douglas Leone, Labor and Workforce Development Agency; Matthew Linzer & Hooman Pejman, University of California Office of the President

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

The meeting opened with Kathy Booth of WestEd providing an update on decisions made by the workgroup at its December meetings, including expanding the governing board to include an additional seat for Labor and Workforce Development Agency, establishing timeline and priorities for phase one of data system development, developing a fiscal estimate for the first year of implementation, and finalizing the legislative report. She noted that the report was submitted to the legislature and the Governor's January budget included $18.8 million in alignment with workgroup recommendations. Finally, she noted that work has begun on a proof of concept for the analytical data system that will link information from CSU, CTC, and CDE on teacher preparation and retention.

Privacy Policy

The subcommittee reviewed a draft privacy policy developed by the Legal Subcommittee that addresses key issues in an FAQ format, such as the components, data providers, and data types in the data system; relevant statutes and resources; and the opt-out process. The document leaves placeholders for

security, data storage, and rules of conduct for staff related to personal information until these items can be finalized in the implementation phase.

Michele Robinson of CDT flagged that the document needs to be reviewed to ensure it aligns with the Information Practices Act (IPA) pertaining to concepts like data quality and integrity, as well as limitations on collection and use. LeAnn Fong-Batkin of WestEd volunteered to compare the draft Privacy Policy to the IPA and to bring the findings back to the Legal Subcommittee.

Rodney Okamoto of CDE asked whether there would be privacy policies established for the operational tools to govern access to student-level information. He recommended that California College Guidance Initiative and eTranscript California policies should be evaluated at the point when the privacy policy gets finalized for the analytical tools.

Baron Rodriguez of WestEd clarified that the IPA is different from the Federal Information Processing (FIPS) standards and noted that there is a difference between data protection and data access policies.

Michele Robinson of CDT shared the National Institution for Standards and Technology (NIST) definition of the fair information practice principles: "principles that are widely accepted in the United States and internationally as a general framework for privacy and that are reflected in various federal and international laws and policies. In a number of organizations, the principles serve as the basis for analyzing privacy risks and determining appropriate mitigation strategies."

PII Definition

The group reviewed a definition of which types of data would be considered PII, which was developed by the Legal Subcommittee, based on the IPA. The subcommittee members had no concerns with proceeding with the recommended scope.

Data Classification Scheme

Kathy Booth of WestEd noted that, given the extensive scope of what would be considered PII, information in the P20W data set would fall into two categories:

- *Public Information* that is not exempt from disclosure, which would be primarily information about institutions (such as address)
- *Personal Information* that must be protected because it identifies or describes an individual, which would include any information about the people in the data system (such as enrollment information)

Baron Rodriguez of WestEd queried the group whether having no distinction between types of information on individuals would be sufficient for agency data protection policies.

Several subcommittee members checked data points of concern against the PII list--such as health information, machine name and IP address, and date of birth—and confirmed they were included.

Lloyd Indig of CHHS noted that as long as the IPA definition aligns with the Health Insurance Portability and Accountability Act (HIPAA), his agency would be comfortable with this approach.

Alan Nakahara of CDE and Gurinder Bains of CSAC indicated that information often gets flagged as PII at the point that it is combined with other data points, but is not considered PII on its own. For example, the K-12 student identifier by itself is not PII, but becomes PII once it is linked to another data point.

However, Michele Robinson of CDT clarified that the recommended definition, which flags each individual item as PII, is correct per the IPA.

Kathy Booth of WestEd noted that in addition to tagging each data point based on whether it is considered public or personal information, each data provider will need to tag each data point for its allowable use, such as whether it can appear in the dashboard, query builder, by request only, or only for the purpose of record matching. These tags would be used to drive the index of data points that are available for data requests. The tagging would also be used to determine when a data point from a specific provider should not be used in the public domain. For example, the Data Definitions Subcommittee has noted that information in the dashboard and query builder tool should designate which individuals are foster youth using CDSS data only. However, educational entities that track foster status could load their information into the Cradle-to-Career Data System and make it available for research purposes.

The subcommittee members indicated they were comfortable moving forward with the proposed data classification categories.

Deidentification Definitions

Baron Rodriguez of WestEd walked subcommittee members through a set of definitions created by the Legal Subcommittee to clarify key terms related to deidentification such as aggregate, summary, and masked. The definitions were based on concepts that had been spelled out in CHHS' deidentification guidelines and by the federal Department of Education. The list is intended to reduce confusion in the context of Cradle-to-Career Data System information.

Subcommittee members asked clarifying questions and Michele Robinson of CDT suggested that the definitions be expanded to clarify how they relate to the data classifications. The subcommittee agreed with moving forward with the definitions, paired with Cradle-to-Career Data System examples.

Deidentification Process

Helen Norris of AICCU described the process used by several subcommittee members to develop a proposed process for deidentifiying information that would appear in the dashboards, query builder, and information released through the expedited data request process. The approach was based on a policy developed by CHHS in alignment with the IPA, and includes recommendations on FERPA compliance released by the National Center for Education Statistics. The group also consulted with Dr. Linette Scott of the Department of Health Care Services who has led the implementation of CHHS' deidentification policy. Finally, Helen Norris noted that the group had focused on balancing techniques that would protect individual identities as well ensure that useful information would still be available to the public. Kathy Booth of WestEd then walked the group through the proposed process.

Akhtar Khan of CDSS affirmed the approach and noted that it had been effective for his agency's research and public dashboards.

Rodney Okamoto of CDE noted that the examples shown in the document had blank cells in the data tables and expressed concern that this would not meet accessibility requirements for screen readers. LeAnn Fong-Batkin of WestEd subsequently checked with the Department of Rehabilitation and noted that they recommend putting text that is the same color as the background into empty cells such as "intentionally blank."

The subcommittee recommended moving the proposed deidentication process forward to the workgroup.

Permission Protocol

Matt Linzer of UC described a protocol for determining access to the data system that was developed by several subcommittee members, in consultation with Bruce Yonehiro, who is a lawyer at CDE. The protocol addresses recommended requirements, applies NIST moderate impact controls, and includes a chart that describes relevant roles. It will need to be modified further by the managing entity during the implementation phase, once the technology tools have been selected, and approved by the governing board. Matt Linzer also clarified that the protocol is designed to accommodate the reality that every data provider will have their own permission protocol, and provides the flexibility to allow for differing agency-level requirements, so long as they meet a common threshold. Finally, the protocol is designed to align with the Interagency Data Exchange Agreement (IDEA), which will provide the legal framework for the Cradle-to-Career Data System.

Daryl Lal of CCCCO thought the chart of roles was robust.

The subcommittee recommended moving the proposed permission protocol forward to the workgroup.

Next Steps

Volunteers were recruited for participating in a homework team to crosswalk key standards including NIST 853, the State Administrative Manual (SAM), and the Statewide Information Management Manual (SIMMS).

Subcommittee members were also alerted that they will need to fill out a survey to clarify which extract-transform-load (ETL) tools they plan to use to load information into the cloud.