Technology & Security Subcommittee Meeting Summary

May 19, 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 use cases, draft Request for Information (RFI), the PowerPoint, and a meeting recording are available at https://cadatasystem.wested.org/meeting-information/technology-security-subcommittee.

The Technology & Security Subcommittee will help develop technology specification requirements to address data structures and privacy considerations. The May 2020 meeting had the following goals:

- Ground the work of this committee by providing an update on the recommended scope for phase one of the California data system
- Recommend how partner entities would provide and store data for the P20W data set and the data request process
- Develop a preliminary tagging and security framework for the P20W data set and the data request process

The following representatives attended the meeting:

Formeka Dent, Antelope Valley Union High School District; Helen Norris, Association of Independent California Colleges and Universities; Jason Piccione, Bureau for Private Postsecondary Education; Barney Gomez & Daryl Lal, California Community College Chancellor's Office; Rodney Okamoto & Alan Nakahara, California Department of Education; Vitaliy Panych & Janet Buehler, California Department of Technology; Adam Dondro, California Health and Human Services Agency; Karissa Vidamo, California Department of Social Services; Ed Sullivan, California State University; Amy Fong & Greg Scull, California School Information Services; Gurinder Bains, California Student Aid Commission; Dan Lamoree, Educational Results Partnership; Todd Ibbotson, Employment Development Department; Douglas Leone, Labor and Workforce Development Agency; Jenni Abbott, Modesto Junior College; and Matthew Linzer & Hooman Pejman, University of California Office of the President

Workgroup Update

The meeting opened with the facilitator providing an update on decisions made by the California Cradle-to-Career Workgroup on their April 29 meeting and answering clarifying questions from subcommittee members.

Guest Speaker: Sean Cottrell, Privacy Technical Assistance Center

Sean Cottrell provided some contextual information about how other states load and store data for intersegmental systems. In response to a question about cloud computing, Sean Cottrell shared the following resource: https://studentprivacy.ed.gov/resources/cloud-computing-faq.

Recommendations on Data Loading and Storage

Next, the participants broke into small groups to develop recommendations for how data could be loaded and stored for the P20W and the Data Request Use Cases. In the subsequent full group discussion, subcommittee members focused on the following concepts:

- If possible, information from the P20W data set should be used to fulfil data requests, rather than having partner entities re-provision this information.
- Even though health and social service data will need to be approved on a case-by-case basis, it will be important to create the infrastructure to transmit those data points when approved. Having technical structures for sharing data will also help with expanding the state data system to include additional sources over time.
- Policies and procedures should reflect the cloud first policy that is already in place for California. Using cloud storage could help to get around constraints associated with a centralized warehouse, because data can be joined across various platforms. The partner entities could load information into cloud repositories associated with the data system, and then control when and how information from those repositories are joined to other data sets. For example, the partner entities could authorize the release of information from these cloud repositories to populate the P20W data set or to fulfil specific data requests, such as for an approved research study or a project initiated by a partner entity.
- The data exchanges available through the Cradle-to-Career system should take into account existing arrangements and requirements, such as fulfilling the federal requirement to reimburse the Employment Development Department for access to wage and earnings data. California can reference guidance on this topic prepared by the federal government at: https://studentprivacy.ed.gov/sites/default/files/resource_document/file/JOINT%20GUIDANCE%20ON%20DATA%20MATCHING%20TO%20FACILITATE%20WIOA%20PERFORMANCE%20REPORTING%20AND%20EVALUATION.pdf
- Clear guidelines are needed to ensure that data remains deidentified when it is shared through public tools, such as the dashboards and query builders, and when it is provided to third parties such as researchers.
- The partner entities have different preferences for how often data should be made available. Several recommended quarterly uploads to ensure data recency. Others felt that data should be provided a few times per year, for the following reasons: information is only provided to state agencies once per term, is provided in different timeframes by different entities, requires time to finalize and validate, and requires partner entities to validate matching each time data are pulled. One participant suggested timeframes should be determined based on the specific data elements required for the use case and their availability.
- While specific file transfer mechanisms were discussed, such as an Application Programming Interface (API) or Secure File Transfer Protocol (SFTP), subcommittee members felt that the technology should be determined later, once options are better understood.

Subcommittee members then crafted the following recommendation about data loading and storage for the Cradle-to-Career Workgroup:

> **P20W Use Case:** Populate a cloud-based common repository that is minimally FedRAMP moderate compliant, with periodic uploads using a file-based template through a secure transfer mechanism.
>
> **Data Request Use Case:** Pull education data first from the P20W data set, then access approved education data not in the P20W data set and the health and social service data through a secure transfer mechanism, and release de-identified unitary data to authorized parties through a self-service mechanism.

In a vote, 13 of the 14 subcommittee members present supported the proposal. Douglas Leone from the Department of Labor voted no because he felt that information should be updated daily.

Recommendations on Tagging and Security

After a presentation from Baron Rodriguez, who previously worked with the federal Privacy Technical Assistance Center, about common security issues for intersegmental data systems, participants returned to small groups to develop recommendations for how sensitive data could be tagged to help keep it secure. In the subsequent full group discussion, subcommittee members focused on the following concepts:

- The partner entities could construct shared categories for classifying data elements—such as personally identifiable information (PII), sensitive data, and information that can be made available for the public—and tag their data before it is loaded, based on their interpretation of which elements belong in each category. Developing these categories may take a significant amount of work because there are multiple state and federal rules related to confidentiality. It will also take work from each partner entity to tag their information.
- The definition for PII could reference policies that have already been established by partner entities or by a variety of international, national, and state regulations. This definition should be determined by the Legal Subcommittee.
- It may be valuable to tag all European students because they are subject to the European Union's General Data Protection Regulation (GDPR). One strategy is to exclude these students from data sets produced by the Cradle-to-Career system.
- While PII definitions and classification categories help with individual elements, some information becomes sensitive when it is combined across data sets. Therefore, an additional level of security may be needed for linked data. Some partner entities, like Health & Human Services, have already developed policies that take into account the sensitivity of combined data.
- Deidentification is another area where policy will be needed, particularly to ensure that each partner entity's guidelines are met.
- Data should be tagged related to ownership to ensure that proper authorities are notified if there is a data breach. For the dashboard and query builder tools, disclosure avoidance strategies should be put in place that meet partner entity requirements.
- Having appropriate permissions will help with managing data security. Policies will be needed to determine who has access to what types of data. For example, the Workgroup will need to determine whether staff at the managing entity could see all types of data.

Subcommittee members then crafted the following recommendation about data tagging and security for the Cradle-to-Career Workgroup:

- Create a set of categories/classifications related to data sensitivity, which each partner can apply to their own data set before it is shared.
- Establish a process for creating the categories/classifications, attending to state, federal, and international requirements, as well as the most restrictive partner entity requirements, to establish the categories. Ensure there is a specific category for data breach notifications. Provide support to partner entities on how to tag their data.

- Develop a review process for deidentification that is specific to the P20W and to the Data Request uses cases. Provide support to partner entities on deidentification process review.
- Implement a role-based access approach, including logging that shows who has accessed which data points and an independent third party that reviews security on an annual basis.

All 14 subcommittee members present voted to support the proposal.

Editing the Request for Information

The facilitation team provided a general update on the work done by the Common Identifier Subcommittee to develop a draft Request for Information (RFI) for a Master Data Management (MDM) solution that will match person records from the partner entities. Several subcommittee members, who serve on both the Technology & Security and Common Identifier Subcommittees, commented that the process had been thorough and collaborative, bringing in perspectives from different types of partner entities. The Technology & Security Subcommittee was tasked with reviewing the document, with an eye to ensuring that any recommendations from the day's meeting were appropriately represented in the document. The subcommittee members elected to do individual reviews of the draft, returning comments no later than the end of the day on Thursday, May 21. They requested that the facilitation team do a final integration of these comments into the document on Friday, May 22. Subcommittee members will then discuss the document with Cradle-to-Career Workgroup members in advance of their May 28 meeting, where the workgroup members will vote on whether the RFI should be released in June. The RFI process will be managed by the California Department of Technology.