Common Identifier/Technology & Security Subcommittee Meeting Summary

September 1, 2020

This document provides a summary of key points that emerged from a half-day 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 September 2020 meeting had the following goals:

- Provide an update on the planning process that may inform recommendations
- Review results from the request for information for a master data management solution
- Develop a recommendation for the workgroup on person-matching, the type of solution, and further information that should be collected this fall

The following representatives attended the meeting:

Common Identifier Subcommittee:

Jonathan Chillas, National University; Todd Hoig, California Community College Chancellor's Office; Ben Allen, Channa Hewawickrama, and Jerry Winkler, California Department of Education; Janet Buehler, California Department of Technology; Jennifer Schwartz and Muree Larson-Bright, California Health and Human Services Agency; Patrick Perry, California Student Aid Commission; Amy Fong and Greg Scull, California School Information Services; Jeff Whitney, California State University; Michele Perrault, Commission on Teacher Credentialing; Akhtar Khan, California Department of Social Services; Amy Faulkner, Labor/Employment Development Department; Eric Goodman, University of California Office of the President; Paco Martorell, University of California, Davis; Dan Lamoree, Educational Results Partnership; John Prindle, University of Southern California; Joseph Hackbarth, California State University, Pomona; Ben Baird, California College Guidance Initiative

Technology & Security Subcommittee:

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; 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; Dan Lamoree, Education Results Partnership; Amy Fong and Greg Scull, California School Information Services; Subash D'Souza and Jeff Whitney, 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

Planning Update

The facilitation team provided an update of decisions made by the Cradle-to-Career Workgroup that could influence the technology needs for the data system and walked participants through an updated chart showing the intake and request process. Participants asked the following clarifying questions:

What is the estimated frequency of source data systems supplying data, and will they only provide certified data? The minimum will be once per year, and only certified data will be uploaded. However, partner entities can upload data more frequently if they choose.

Will requestors be able to download data in a file? Based on decisions to date, requestors would remotely access data in a secure enclave and would not be able to remove information from this location. However, this recommendation has not been yet formally reviewed and endorsed by the Technology & Security Subcommittee.

Will EDD match employment records for each request or do a comprehensive match once a year? How will they do the match? They would do a yearly update, using deduplicated social security numbers from the master data management (MDM) solution. From the resulting matched information, EDD would remove social security numbers and use the unique identifier assigned by the MDM to allow employment information to be linked to the unitary records in the P20W data set. This arrangement would not preempt any agreements between individual data providers and EDD for other purposes. Because EDD matches using social security numbers, the P20W data set is unlikely to have employment records for individuals who did not enter a postsecondary institution (the most reliable source of social security numbers among the data contributors), and the system will not have information on those who decline to share their social security numbers or on undocumented individuals.

One participant asked that a second version of this chart be created that is easier for non-technical people to understand, particularly regarding how information would be linked between various data sources. The facilitation team noted that it might be helpful to do some more detailed charts as well that cover how the MDM works related to matching and indexing, as well as doing separate charts for the technical flow and for the legal process.

Results of the Request for Information (RFI)

Katie DeAngelis of CDT described the process used to create the RFI, solicit information about possible MDM solutions, and compile results about the types of options that might be available. She reminded the group that an RFI is a market research activity, not an effort to evaluate individual vendors. Because the technical specifications for the data system had not been clearly determined at the point the RFI was released, the range of responses was quite varied, making it difficult to do an apples-to-apples comparison. This is particularly true for the budgets provided, some of which were only for core MDM features and others which included a broader range of related services such as data warehousing and visualization. Finally, she noted that the RFI had yielded an unusually large number of responses.

Erin Carter of WestEd provided an overview of the responses. Responses were reviewed in four ways: 1) ability to provide minimum security and compliance features, 2) ability to provide minimum MDM capabilities, 3) provider experience level, and 4) ability to address various types of feature requirements including the system solution, data management, data processing, data matching, and optional items on data publishing and release.

Vendors described four primary types of solutions:

- Commercial off the shelf MDM solutions
- Full-service solutions with a proprietary or partnered MDM
- Custom development solutions
- Other software solutions that did not provide MDM functionality

Many of the respondents described relevant experience, including supporting one of the partner entities, implementing a longitudinal data system, working with education data, and working with health and human services data.

Projected costs varied significantly, and some respondents did not provide budget information. Among responses received, average costs for the start-up phase ranged from $2.6 million to $3.8 million, with annual costs after the start up phase averaging between $1 million and $1.4 million. Respondents reported an average of 20 months for full implementation.

The subcommittee members who read the responses recommended that the following items be considered when developing a request for proposals (RFP):

- Budget and timeline
- Staffing level and decision-making authority for the managing entity
- More granular project scope and requirements
- Clarifying how the solution would integrate with and leverage state data systems and solutions

The group asked clarifying questions of the readers:

What types of additional functionality beyond MDM did respondents describe, and where these integrated with the MDM solution or modular options? In most cases the additional components were integrated and included resources such as data warehouses, research enclaves, and business intelligence tools. Some respondents also described implementation support related to data governance and stewardship, or managing the development process.

Would it be possible to receive the names of the RFI respondents? These are available through a public records act release. Submit requests at this link: https://cdt.ca.gov/public-records-request/

What did respondents indicate would be delivered during the start-up phase? This was variable, depending on the solution being proposed. Most comprehensive providers described iterative releases starting with a minimum viable product.

Did the respondents that provide public-facing tools offer custom solutions or did they partner with other vendors? Several offered integrated solutions while others recommended working with common commercial visualization tools.

Recommendations Regarding Person-Matching and Technology Solutions

Whether to recommend an MDM solution

Although a concern was expressed that not all RFI respondents were able to meet the requirements, the group ultimately noted that there were sufficient qualified respondents to indicate that an MDM solution is a feasible approach for person-matching and developing unique identifiers for the Cradle-to-

Career system. In each of the review rubrics, there were ten or more respondents that met the criteria, with seven respondents that met all criteria. At the point that a more detailed features list can be developed, it is likely that respondents could be more specific about how they would support the desired functionality.

Whether to recommend an MDM-only or a more comprehensive solution

Given that a number of the respondents offered comprehensive solutions, the group discussed whether it would be preferable to put out a single RFP for all desired technical functionality or take a more modular approach.

One subcommittee member was concerned that if the RFP only seeks an off-the-shelf MDM, it may be too rigid. Another participant indicated it would be valuable to work with a vendor that is flexible about technology solutions and has access to many different types of products. Having a single provider reduces the workload on the managing entity because the service provider would take on the task of coordinating across the modules. This can also help avoid a situation where multiple vendors are unable to coordinate to provide the desired solution. However, another subcommittee member stated that one value of outsourcing is the ability to bring in just the expertise that is needed at a given time.

The group discussed the importance of being able to balance the expected responsibilities between the managing entity staff and any vendor(s) selected. Some participants flagged the challenges of hiring contractors in a union setting, while others worried that the state would not be able to hire qualified IT staff. One participant requested examples of how state agencies have balanced responsibilities between staff and technology vendors. GovOps' current work with vendors to develop a homeless data system was flagged as an example of a flexible solution that was leveraging the skills of highly qualified staff, while CDE's experience developing CALPADS with an outside vendor and little staff support at the outset was offered as a cautionary tale.

One subcommittee member noted that when working with a vendor, it is critical to ensure knowledge transfer to staff. This is particularly true in tight budget years, when funding for external sources may be cut, or in cases where vendors demand high fees or decide to stop supporting a product. Another participant emphasized the importance of having core technical staff in place at the beginning of the project. The group was supportive of the idea that the contract should require the vendor to train the managing entity to be able to manage the product independently.

Some workgroup members noted that other states, such as Kentucky, had built custom solutions, and that California could use expertise from partner entities such as CHHS to build a model that meets California's needs. Baron Rodriguez of WestEd, who previously provided technical assistance to state longitudinal data system, noted that the states that built custom solutions had much smaller populations and fewer agencies involved.

Baron Rodriguez also clarified that several state systems that had relied exclusively on outside vendors to deliver technology services had failed. A hybrid model is more likely to be successful and can also be more cost effective. Other subcommittee members, reflecting on their own experience, echoed the recommendation of taking a hybrid approach—just as it is risky to outsource expertise, it is also difficult to build everything in-house. Many members of the group indicated that a hybrid solution that partners strong technical staffing with external vendors would be the best approach.

An informal poll regarding whether it was better to solicit a single vendor that could deliver a comprehensive solution or for the managing entity to select a variety of solutions showed that the majority supported taking a flexible approach. Further discussion surfaced the idea that flexibility might include one vendor that coordinates multiple products and partners. The group also discussed the risk of working with multiple vendors independently, because the contractors may not work together effectively or there could be poor communication and overlap between staff-based and contractor-based projects. The group stressed the importance of the managing entity staff being able to oversee all of the solutions, verify that they are delivering the expected functionality, and develop the skills to maintain and potentially adapt the product over time.

Gathering additional information about existing state technology solutions

Given CHHS' interest in leveraging its person-matching, open data portal, and secure data enclave solutions to support the Cradle-to-Career system, as well as other efforts underway such as the homeless data system and a new early care and learning data system, the group discussed what additional information would be helpful to gather about existing state technology solutions.

One participant expressed concern that existing systems may have been tailored for other requirements that may not meet the needs of the Cradle-to-Career system. Another worried that redirecting an existing system to a new purpose might dilute the ability of that system to serve its intended purpose. A third indicated that information should only be gathered from systems that were interested in supporting the Cradle-to-Career system.

One participant suggested getting more information about open data platforms and how those data standards could support the development process. A number of participations indicated that it would be helpful to have product demonstrations. However, others noted that it would be better to wait until the RFP was under development and the specifications list was more clear. Katie DeAngelis of CDT reminded the group that if they have any product demos, then all interested vendors must also be given a chance to demonstrate their products.

VOTES

The group voted to forward the following recommendations to the workgroup:

- Given that there are a number of qualified providers of MDM solutions, the Cradle-to-Career system should pursue this type of approach for person-matching.
- When soliciting proposals for a technical solution, allow providers to either provide a comprehensive solution or propose modular solutions.
- The managing entity should hire technical staff who can oversee the selection and implementation of the technical solutions, have sufficient expertise to knowledgably manage the technical solutions, and potentially develop custom components as needed.
- An assessment should be conducted to better understand ways that commercial solutions could be integrated with existing and emerging state technologies.