Legal and Technical Framework for the Cradle-to-Career Data System

Note: this document has been amended from the version approved at the June 2020 workgroup meeting to reflect subsequent decisions and shifts in terminology.

Overall Framework

- *Legal Framework:* Each data provider retains control over their own information. Data provider participation in the state data system and requirements for the managing entity will be defined through a **Participation Agreement.** Data providers and institutional members of the governing board can authorize data sharing with each other using either the **Interagency Data Exchange Agreement (IDEA)**, which governs data use for members of the executive branch, or the **Master Data Exchange Agreement (MDEA),** which is specific to the Cradle-to-Career Data System. IDEA and MDEA will be paired with a **Business Use Case Proposal** (BUCP) that outlines the specifics parameters for each data access instance. A **library of legal agreements**, based on the relevant federal framework(s) that allow for data access (education, health, and financial aid) will be used for third parties that are approved to access data through an authorized **Data Request Process**.
- *Provisioning:* Each data provider will upload a subset of their information into individualized cloud-based repositories associated with the state data system, using a secure transfer mechanism. Additional data sets needed to fulfil approved studies, such as National Student Clearinghouse files or identifiers for students related to an evaluation study, will also be loaded to individualized cloud-based repositories. The Master Data Management (MDM) solution will be the mechanism by which person matching is conducted and data points are combined to create data sets. The managing entity will be responsible for creating the data sets in accordance with specific BUCPs or third party legal agreements.
- *Permissions:* A **Permission Protocol** will be used to allow each data provider to designate who within their organization has access to the information in the cloud and the secure data enclave. The managing entity will also have access to this information to support state data system activities.
- *Security:* Security will be structured in compliance with a **Data Security Policy** and a **Personally Identifiable Information (PII) Definition** that is part of the California Cradle-to-Career Data System governance framework. The data providers will tag information that they upload to the cloud using a **Data Classification Scheme** to indicate how sensitive each data point is, including whether the data element is PII and whether the information can be included in the P20W Data Set or for approved third party requests. Each data provider will determine which data elements from their data set are assigned to each data classification

category. When data sets are created, the MDM will tag each element to clarify the associated data provider, to support notification in the instance of a breach.

- *Deidentification*: Disclosure avoidance will be structured in compliance with a **Data Suppression Protocol** that is part of the California Cradle-to-Career Data System governance framework, which will reference the Data Classification Scheme.
- *Timing:* The timing of the data uploads to the cloud will be determined by each data provider but will be no less than once per year for person matching and the P20W data set, and upon approval for data requests. So, for example, the University of California Office of the President could elect to upload data at the end of each term to minimize the work needed to fulfil data requests, while the California Department of Social Services agency could upload only the information necessary for person matching and the data points identified for the P20W data set and include other items only at the point of request.
- *Matching*: The person matching index will be updated once per year, when the P20W data set is rebuilt.
- *Funding*: If fees are required for data requests, they will be structured in compliance with a **Payment Policy** that is part of the California Cradle-to-Career Data System governance framework, which will be aligned with any federal and state requirements.
- *Changes to Scope*: Over time, the governing board can recommend additional data sets that should be created or maintained by the managing entity.

Note: in the examples below, color coding helps to clarify the level of deidentification. The characteristics noted in colored text in this box are also noted as the first bullet for each item below.

- Green indicates deidentified aggregate data, where PII has been removed and the suppression protocol has been applied.
- Blue indicates anonymized individual-level data, where personal identifiers have been removed but information is unsuppressed.
- Red indicates identifiable individual-level data.

P20W Data Set

- *Level of Deidentification:* Anonymized individual-level data, where personal identifiers have been removed but information is unsuppressed
- *Participating Data Providers*: Bureau for Private Postsecondary Education, California Community College Chancellor's Office, California Department of Education, California Student Aid Commission, California State University, Department of Apprenticeship Standards, Department of Health Care Services, Department of Social Services, Employment Development Department, University of California Office of the President, and independent colleges

- *Legal Framework:* The data providers contributing to the P20W data set will sign the Participation Agreement that governs the content and specifies the purpose of populating the dashboards and query builder. This agreement will delegate authority to the managing entity to manage the data set.
- *Provisioning, Matching & Technology:* Once per year, to create the P20W data set, the MDM solution will refresh the person match index and pull the designated elements in the cloud repository associated with each participating data provider. Each unique individual will be assigned an identifier that is specific to that data pull.
- *Access & Permissions*: The P20W data set will only be accessed by the managing entity. The P20W data set will be used to populate the dashboards and query builder, which will be publicly available. The dashboards and query builder will apply a suppression protocol so that the information displayed will be deidentified aggregate data.

IDEA/MDEA Signatory PII Data Requests

- *Level of Deidentification:* Identifiable individual-level data
- *Participating Data Providers*: Optional
- *Legal Framework:* IDEA/MDEA signatories will sign a BUCP that governs the purpose, content, and upload timing. This agreement will delegate authority to recipient to conduct the analysis and to the managing entity to manage the data set.
- *Provisioning, Matching & Technology:* To create the data set, the MDM solution will pull the designated elements in the cloud associated with each participating data provider. If the requested data point is not already in the cloud, the data provider will upload it. Each unique individual will be assigned an identifier that is specific to that data pull.
- *Access & Permissions*: The requestor will be allowed to download the data set for the approved use. The data set will be made available to authorized staff of the managing entity and the data providers that provided information. Or, the requestor will be able to conduct analyses in a secure data enclave, with the data providers and the managing entity having access to the underlying data set. Each organization will determine which of their staff has permission to access the secure data enclave.

IDEA/MEDA Signatory Anonymized Individual-Level Data Requests

- *Level of Deidentification:* Anonymized individual-level data, where personal identifiers have been removed but information is unsuppressed
- *Participating Data Providers*: Optional
- *Legal Framework:* IDEA/MDEA signatories will sign a BUCP that governs the purpose, content, and upload timing. This agreement will delegate authority to the managing entity to manage the data set.
- *Provisioning, Matching & Technology:* To create the data set, the MDM solution will pull the designated elements in the cloud repository associated with each

participating data provider. If the requested data point is not already in the cloud, the data provider will upload it. Each unique individual will be assigned an identifier that is specific to that data pull.

- *Access & Permissions*: The requestor, contributing data providers, and managing entity will be allowed to download the data set for the approved use. Or, the requestor will be able to conduct analyses in a secure data enclave, with the data providers and the managing entity having access to the underlying data set. Each organization will determine which of their staff has permission to access the secure data enclave.

Third-Party Data Requests

- *Level of Deidentification:* Could be either 1) deidentified aggregate data, where PII has been removed and disclosure rules have been applied or 2) anonymized individual-level data, where personal identifiers have been removed but information is unsuppressed. For anonymized individual-level data, third parties will be required to provide proof of training on data protections and sign non-disclosure agreements prior to accessing this data set.
- *Participating Data Providers*: Optional
- *Legal Framework:* Once the data request has been approved through the Data Request Process, data providers contributing to the data set will sign a legal agreement based on the appropriate templates in the legal agreement library. Legal agreement signatories will include the data providers, the managing entity, and the data requestor. The legal agreement will specify purpose, content, and upload timing. It will also delegate authority to the managing entity to manage the data set.
- *Provisioning, Matching & Technology:* To create the data set, the MDM solution will refresh the person match index and pull the designated elements in the cloud associated with each participating data provider. If the requested data point is not already in the cloud, the data provider will upload it. Each unique individual will be assigned an identifier that is specific to that data pull.
- *Access & Permissions*: The requestor will be able to conduct analyses in a secure data enclave, with the data providers and the managing entity having access to the underlying data set. The requestor will need to specify which of their staff has permission to access the secure data enclave.