

Cradle-to-Career Data and Tools Advisory Board Staff Report

Date Report Issued: September 23, 2025
Attention: Members of Data and Tools Advisory Board
Subject: **Update on the Research Request Process, Phase One of the Query Builder, and the Secure Data Enclave (Agenda Item Five)**
Staff Contact: Ryan Estrellado, Director of Data Programs

At this meeting, the Data and Tools Advisory Board (DTAB) will have the opportunity to receive updates on Cradle-to-Career Data System's (C2C) research request process, phase one of the query builder, and the secure data enclave. This staff report is broken down into three main parts: the research request process, section one, query builder, section two, and secure data enclave, section three.

Requested Action:

There is no requested action for this item. This is an informational item only.

Background:

[Education Code Section 10866\(b\)\(6\)](#) charges the Governing Board with "creating, and revising from time to time, in consultation with the advisory boards, a data request process for use by researchers, policymakers, education systems, schoolsites, and college campuses ...".

Section One: Research Request Process

The Research Request Process employs a three-tiered system to facilitate data requests that enable the California public to use data in the system to conduct research that addresses disparities in opportunities in student outcomes. The three tiers are the Query Builder, Expedited Data Requests, and Comprehensive Data Requests.

Section One, Part A: Office Update on the Research Request Process

The August 15 Governing Board (Board) meeting was used as an opportunity to gather initial feedback from the Board on critical design elements of the Research Request Process forms. Specifically, the Office shared forms for receiving research requests, with the goal of making the Research Request Process accessible and transparent ([Education Code Section 10867\(b\)\(7\)\(H\)](#)). These forms can be found in Attachment A. Additionally, the Office shared its current thinking on Open Science best practices and how these can be applied to C2C's Research Request Process.

Section One, Part B: Member Questions - Research Request Process

During the DTAB meeting, Advisory Board members will have the opportunity to share written and verbal input on these in-depth questions:

1. Reflecting on what you have heard so far, what clarifying questions do you have about the forms found in Attachment A?
2. How does what you have heard so far about the research request process compare to what you have seen in the field?
3. What are the risks we should be aware of as we develop this product?

Section Two: Query Builder

[Education Code Section 10867\(b\)\(4\)\(B\)](#) charges the Office to implement public educator- and researcher-facing tools including a query builder. The Office is focusing on the first phase of the query builder, with plans to receive user feedback to inform regular expansion and development of the tool.

Section Two, Part A: Office Update on the Query Builder

The Office has developed a new query builder tool to expand access to the data behind the Student Pathways Data Story. The first phase provides a clear, well-labeled web page where researchers and other users can download data files directly. This initial version will prioritize usability, accessibility, and alignment with the existing C2C site. While future enhancements will introduce more interactivity and customization, this first step lays the groundwork for broader access and engagement. The Office implemented feedback from previous

DTAB and CEAB sessions on the first iteration of the query builder and released it to the public on September 16, 2025.

Section Two, Part B: Member Questions - Query Builder

During the DTAB meeting, Advisory Board members will have the opportunity to share written and verbal input on these in-depth questions:

1. Reflecting on what you have heard so far, what clarifying questions do you have about the first phase of the query builder?
2. What are the risks we should be aware of as we develop this product?

Part Three: Secure Data Enclave

As part of the research request process, the secure data enclave will provide researchers with a secure compute environment to analyze deidentified individual level data for approved research requests.

Section Three, Part A: Office Update on the Secure Data Enclave:

The Data Programs team and the Data Infrastructure team are working to define the initial version of the secure data enclave. The secure data enclave will provide a secure computing environment for users with approved research requests to conduct research. Currently, the Office is defining the scope of the data tools for researchers to use in the secure data enclave, including statistical programming tools. The Office is working on identifying a set of candidate users to pilot the secure data enclave.

Section Three, Part B: Member Questions - Secure Data Enclave

During the DTAB meeting, Advisory Board members will have the opportunity to share verbal input on this in-depth question:

1. What are the risks we should be aware of as we develop the secure data enclave?

Research Request Intake Form

Complete this form to begin the following types of data requests from the California Cradle- to-Career (C2C) Data System:

- ❖ **Expedited Data Request:** requests for deidentified aggregate data that is not available through the query builder. These requests will be reviewed by C2C staff and shared with Data Providers before being fulfilled. These requests do not require a legal agreement.
- ❖ **Comprehensive Data Request:** requests for individual-level data or to match a cohort of individuals against C2C data. These requests will be reviewed by C2C staff and approved by Data Providers and the Commission for the Protection of Human Subjects. If approved, requestors will sign a legal agreement and provide proof of research training (e.g., Collaborative Institutional Training Initiative (CITI) Program). C2C staff and Data Providers will review analytical outputs to ensure data privacy.

Please refer to C2C's list of available data points as described [here](#).

Requestor Information

1. Requestor Name
2. Requestor Organization(s)
 - ❖ What type of organization is requesting data? Check all that apply.
 - Advocacy Organization
 - Community Organization
 - Research Organization
 - Media
 - Nonprofit-NGO
 - Local Education Agency
 - Local Government Agency
 - State Agency
 - Researcher
 - Legislative
 - Policymaker
 - Postsecondary Institution/System
3. Requestor Emails
4. Requestor Phone Numbers
5. Is this request on behalf of your employer, a Data Provider, or for an independent project?
6. When do you need the requested data?
7. Is this an update to an earlier request?
8. Prior request number (if none, leave blank)
9. Will you be providing C2C with any other data to link or combine with your requested data? If "Yes", please also fill out the data submission request form.

Project Information

10. What is the title of your project?
 11. What type of project will be supported by this data request?
Check all that apply.
 - ❖ California state legislative report
 - ❖ Federal reporting requirements
 - ❖ Grant-funded research
 - ❖ Thesis or dissertation
 - ❖ Other report by state or local government agency
 - ❖ Data to inform organizational decision-making
 - ❖ Other (Please Specify):
 12. Provide an executive-summary describing the request and its purpose (no more than 500 words).
 13. What questions will you answer with the requested data?
 14. What aspect of the Cradle-to-Career Data system are you interested in studying? Check all that apply.
 - ❖ Early Childhood to Pre-K
 - ❖ K-5
 - ❖ Middle School
 - ❖ High School
 - ❖ Postsecondary
 - ❖ Workforce
 - ❖ Other
 15. Describe your population of interest. Are you interested in a snapshot or cohort? What individuals should be included? Please be as conservative as possible and request the minimal amount of data to facilitate your request. Overly broad requests will be denied.
 16. What timeframe should C2C use to identify your group of interest? List the first and last years.
 17. Have you consulted with any of C2C's Data Providers about your request?
 18. Have you consulted with any other state or local entities regarding your request?
-

19. If you have consulted with C2C's Data Providers or other state/local entities, please elaborate on your consultation.
20. Is this a one-time request or an ongoing request for data? If on-going, please specify how often you would expect to receive data.
21. In what format should the data be delivered? When possible, C2C will try to accommodate specifications.

Open Science Practices

The Office is committed to Transparency and Openness Promotion Guidelines (TOPS) ([TOP Guidelines](#)) C2C is interested in how its data is being used to explore different scientific topics. By answering these questions before receiving data, the Office will create a record of researchers hypotheses at a point-in time. The questions below request information on the intended purposes, methods of analysis, and expected outcomes related to the proposed use of C2C data. Note that C2C expects that requestors use data requested via this form only for the purposes specified in the research project requestors detail in this section.

22. Has this project been registered/pre-registered with the Center for Open Science, Dataverse, or other registry? If "Yes", please provide a link to the existing registration.
23. Will you commit to pre-registering this work and your planned analysis before receiving the data?
 - Yes, I plan to pre-register this work.
 - No, this work is exploratory or serves another purpose.
24. Please provide any existing citations that provide a basis for your research.
25. For each of the following, please indicate whether, prior to accessing C2C data, you intend to (1) disclose that this information will be available, (2) require the information to be available, or (3) arrange for an independent third party to verify your protocol.
 - Proposed study protocol
 - Proposed analysis plan
 - Plan for transparency of:
 - o analytical materials
 - o data
 - o analytical code
 - Reporting plans and guidelines
26. Prior to accessing C2C data, have you arranged for an independent party to:

- Review your computational processes for reproducibility?
 - Review results to ensure information is accurately reported?
27. Is your proposed research study a verification of an existing research project?
28. If “Yes” to the above question, please select the type of verification study:
- Replication: replication of existing study
 - Registered report: registered study that has been pre-accepted for publication
 - Multiverse: multi-process approach with the same data
 - Many analysts: multi-team analysis of the same data

Compliance/Institutional Review Boards

29. Have you submitted an application to the relevant Institutional Review Board, and if so, has your request been approved?
 - Which IRB have you submitted to?
 - o CPHS
 - o Your own institution
30. If your request has been approved, has your research been categorized as exempt review, expedited review, full review, or not human subjects research?
 - If exempt or not human subjects research, please describe the reason for this determination (ex. secondary research where identifiable information cannot be linked back to an individual)
31. Do you agree to follow C2C's data security practices and guidelines for researchers?
32. How long will C2C need to retain the data for your request?
33. Do you certify that the elements you are requesting are the minimal data necessary to conduct the analysis?
34. Does your intended analysis focus on any of the following special populations?
 - Children/minors
 - Incarcerated individuals
 - Individuals with cognitive impairments
 - Economically/educationally disadvantaged individuals
 - Pregnant individuals

Data Sharing and Reuse

35. Once this data is generated, are there any reasons why this request cannot be shared? (C2C's policy will be such that all aggregated data requests will be posted publicly, along with the underlying methodology and code subject to privacy and security policy). For comprehensive data requests, the underlying code must be placed in a publicly available repository or registry.
36. If you are requesting to embargo the data request from the public, please provide a rationale and the duration of the embargo.

Requested Data Points

37. Please Identify the Data Providers you are requesting data from. Check all that apply.

BPPE	Bureau for Private Postsecondary Education
CCC	California Community Colleges
CDE (K12)	California Department of Education
CSAC	California Student Aid Commission
CSU	California State University
CTC	Commission on Teacher Credentialing
CWDB	California Workforce Development Board
DDS	Department of Developmental Services
DHCS	California Department of Health Care Services
DIR	Department of Industrial Relations
DSS	Department of Social Services
EDD	Employment Development Department
CDE (ELC)	Early Learning and Care
ETP	Employment Training Panel
IND	Independent Institutions of Higher Education
SIB	ScholarShare Investment Board
UC	University of California

38. Please identify the specific data points you are requesting access to from C2C's data dictionary [here](#).

39. Please select the type of data request you are initiating:

- **Expedited Data Request:** requests for deidentified aggregate data that is not available through the query builder. *[Skip to Expedited Data Request Section]*
- **Comprehensive Data Request:** requests for unitary data or to match a cohort of individuals against C2C data. *[Skip to Comprehensive Data Request Section]*

Expedited Data Request

Expedited Data Request: requests for deidentified aggregate data that is not available through the query builder. These requests will be reviewed by C2C staff and by Data Providers before being fulfilled. These requests do not require a legal agreement.

40. Please confirm that the aggregate data you are requesting is not available in the query builder.
41. For each data point please describe the statistics you want calculated and how you would like the results aggregated (i.e., mean college-going rate by socioeconomic status and gender). Please note all results will be subject to C2C's review processes.

Supplementary Questions

These questions will help provide transparency into how the P20W data is being used as well as the types of individuals who are using the data system.

Answering these questions accurately will aid public reporting about the use of the data system. Please answer the following questions to the best of your ability.

42. Which of the following best describes your work? Check all that apply.

- Descriptive
- Predictive Analysis
- Casual Analysis

43. Does this work focus on any of the following subpopulations?

- Gender
- Race
- Individuals with disabilities
- Socio-economic status
- English Language Learners

44. Does this work focus on specific geographical regions?

45. Does this work focus on specific institutions (i.e., a local school district)?

46. If your request was not able to be fulfilled, please describe your most likely course of action:

- Do nothing
- Conduct the Analysis using aggregated data
- Seek out data directly from a state agency
- Seek out data directly from a local agency
- Conduct the analysis using another resource (describe)
- Other: Please Explain

Research Request Intake Form

Complete this form to begin the following types of data requests from the California Cradle- to-Career (C2C) Data System:

- ❖ **Expedited Data Request:** requests for deidentified aggregate data that is not available through the query builder. These requests will be reviewed by C2C staff and shared with Data Providers before being fulfilled. These requests do not require a legal agreement.
- ❖ **Comprehensive Data Request:** requests for individual-level data or to match a cohort of individuals against C2C data. These requests will be reviewed by C2C staff and approved by Data Providers and the Commission for the Protection of Human Subjects. If approved, requestors will sign a legal agreement and provide proof of research training (e.g., Collaborative Institutional Training Initiative (CITI) Program). C2C staff and Data Providers will review analytical outputs to ensure data privacy.

Please refer to C2C's list of available data points as described [here](#).

Requestor Information

1. Requestor Name
2. Requestor Organization(s)
 - ❖ What type of organization is requesting data? Check all that apply.
 - Advocacy Organization
 - Community Organization
 - Research Organization
 - Media
 - Nonprofit-NGO
 - Local Education Agency
 - Local Government Agency
 - State Agency
 - Researcher
 - Legislative
 - Policymaker
 - Postsecondary Institution/System
3. Requestor Emails
4. Requestor Phone Numbers
5. Is this request on behalf of your employer, a Data Provider, or for an independent project?
6. When do you need the requested data?
7. Is this an update to an earlier request?
8. Prior request number (if none, leave blank)
9. Will you be providing C2C with any other data to link or combine with your requested data? If "Yes", please also fill out the data submission request form.

Project Information

10. What is the title of your project?
 11. What type of project will be supported by this data request?
Check all that apply.
 - ❖ California state legislative report
 - ❖ Federal reporting requirements
 - ❖ Grant-funded research
 - ❖ Thesis or dissertation
 - ❖ Other report by state or local government agency
 - ❖ Data to inform organizational decision-making
 - ❖ Other (Please Specify):
 12. Provide an executive-summary describing the request and its purpose (no more than 500 words).
 13. What questions will you answer with the requested data?
 14. What aspect of the Cradle-to-Career Data system are you interested in studying? Check all that apply.
 - ❖ Early Childhood to Pre-K
 - ❖ K-5
 - ❖ Middle School
 - ❖ High School
 - ❖ Postsecondary
 - ❖ Workforce
 - ❖ Other
 15. Describe your population of interest. Are you interested in a snapshot or cohort? What individuals should be included? Please be as conservative as possible and request the minimal amount of data to facilitate your request. Overly broad requests will be denied.
 16. What timeframe should C2C use to identify your group of interest? List the first and last years.
 17. Have you consulted with any of C2C's Data Providers about your request?
 18. Have you consulted with any other state or local entities regarding your request?
-

19. If you have consulted with C2C's Data Providers or other state/local entities, please elaborate on your consultation.
20. Is this a one-time request or an ongoing request for data? If on-going, please specify how often you would expect to receive data.
21. In what format should the data be delivered? When possible, C2C will try to accommodate specifications.

Open Science Practices

The Office is committed to Transparency and Openness Promotion Guidelines (TOPS) ([TOP Guidelines](#)) C2C is interested in how its data is being used to explore different scientific topics. By answering these questions before receiving data, the Office will create a record of researchers hypotheses at a point-in time. The questions below request information on the intended purposes, methods of analysis, and expected outcomes related to the proposed use of C2C data. Note that C2C expects that requestors use data requested via this form only for the purposes specified in the research project requestors detail in this section.

22. Has this project been registered/pre-registered with the Center for Open Science, Dataverse, or other registry? If "Yes", please provide a link to the existing registration.
23. Will you commit to pre-registering this work and your planned analysis before receiving the data?
 - Yes, I plan to pre-register this work.
 - No, this work is exploratory or serves another purpose.
24. Please provide any existing citations that provide a basis for your research.
25. For each of the following, please indicate whether, prior to accessing C2C data, you intend to (1) disclose that this information will be available, (2) require the information to be available, or (3) arrange for an independent third party to verify your protocol.
 - Proposed study protocol
 - Proposed analysis plan
 - Plan for transparency of:
 - o analytical materials
 - o data
 - o analytical code
 - Reporting plans and guidelines

26. Prior to accessing C2C data, have you arranged for an independent party to:
- Review your computational processes for reproducibility?
 - Review results to ensure information is accurately reported?
27. Is your proposed research study a verification of an existing research project?
28. If “Yes” to the above question, please select the type of verification study:
- Replication: replication of existing study
 - Registered report: registered study that has been pre-accepted for publication
 - Multiverse: multi-process approach with the same data
 - Many analysts: multi-team analysis of the same data

Compliance/Institutional Review Boards

29. Have you submitted an application to the relevant Institutional Review Board, and if so, has your request been approved?
 - Which IRB have you submitted to?
 - o CPHS
 - o Your own institution
30. If your request has been approved, has your research been categorized as exempt review, expedited review, full review, or not human subjects research?
 - If exempt or not human subjects research, please describe the reason for this determination (ex. secondary research where identifiable information cannot be linked back to an individual)
31. What procedures do you have in place for data security practices?
32. What is your plan for destruction of individual level data once your project has been completed?
33. Do you certify that the elements you are requesting are the minimal data necessary to conduct the analysis?
34. Does your intended analysis focus on any of the following special populations?
 - Children/minors
 - Incarcerated individuals
 - Individuals with cognitive impairments
 - Economically/educationally disadvantaged individuals
 - Pregnant individuals

Data Sharing and Reuse

35. Once this data is generated, are there any reasons why this request cannot be shared? (C2C's policy will be such that all aggregated data requests will be posted publicly, along with the underlying methodology and code subject to privacy and security policy). For comprehensive data requests, the underlying code must be placed in a publicly available repository or registry.
36. If you are requesting to embargo the data request from the public, please provide a rationale and the duration of the embargo.

Requested Data Points

37. Please Identify the Data Providers you are requesting data from. Check all that apply.

BPPE	Bureau for Private Postsecondary Education
CCC	California Community Colleges
CDE (K12)	California Department of Education
CSAC	California Student Aid Commission
CSU	California State University
CTC	Commission on Teacher Credentialing
CWDB	California Workforce Development Board
DDS	Department of Developmental Services
DHCS	California Department of Health Care Services
DIR	Department of Industrial Relations
DSS	Department of Social Services
EDD	Employment Development Department
CDE (ELC)	Early Learning and Care
ETP	Employment Training Panel
IND	Independent Institutions of Higher Education
SIB	ScholarShare Investment Board
UC	University of California

38. Please identify the specific data points you are requesting access to from C2C's data dictionary [here](#).
39. Please select the type of data request you are initiating:
- **Expedited Data Request:** requests for deidentified aggregate data that is not available through the query builder. *[Skip to Expedited Data Request Section]*
 - **Comprehensive Data Request:** requests for unitary data or to match a cohort of individuals against C2C data. *[Skip to Comprehensive Data Request Section]*

Comprehensive Data Request

Comprehensive Data Request: requests for unitary data or to match a cohort of individuals against C2C data. These requests will be reviewed by C2C staff and evaluated by Data Providers and the Commission for the Protection of Human Subjects. If approved, requestors will sign a legal agreement and provide proof of research training (e.g., Collaborative Institutional Training Initiative (CITI) Program). Data will be deidentified and provisioned in a secure data enclave. C2C staff will review analytical outputs to ensure data privacy and Data Providers will review reports for inclusion in the C2C Research Library.

40. Have you received outside funding to conduct this study?
41. If "Yes" to the above question, please explain what funding sources support this study.

Supplementary Questions

These questions will help provide transparency into how the P20W data is being used as well as the types of individuals who are using the data system.

Answering these questions accurately will aid public reporting about the use of the data system. Please answer the following questions to the best of your ability.

42. Which of the following best describes your work? Check all that apply.

- Descriptive
- Predictive Analysis
- Casual Analysis

43. Does this work focus on any of the following subpopulations?

- Gender
- Race
- Individuals with disabilities
- Socio-economic status
- English Language Learners

44. Does this work focus on specific geographical regions?

45. Does this work focus on specific institutions (i.e., a local school district)?

46. If your request was not able to be fulfilled, please describe your most likely course of action:

- Do nothing
- Conduct the Analysis using aggregated data
- Seek out data directly from a state agency
- Seek out data directly from a local agency
- Conduct the analysis using another resource (describe)
- Other: Please Explain

C2C Data Submission Form Outline

If you would like to send C2C data to merge with the data you have requested, please answer the following questions about the types of data you would like to match. The information provided below will be included in any relevant legal agreements established between C2C and the requestor.

Dataset Characteristics

1. Please provide an existing C2C Data Request form submission number.
 - a. Is this Data Request number for individual data or for aggregated data? Please note that the final matched dataset will match the format indicated in the original Data Request.
2. Are the elements in your dataset reported at the individual level or the aggregate level?
3. How many records require matching?
4. Please provide a list of data fields, files, and descriptions for the files you would like to provide. Please classify the fields confidentiality according to the FIPS 199, Standards for Security Categorization of Federal Information and Information Systems. Please also indicate which data fields contain personally identifiable information (PII) or protected health information (PHI), sensitive information (ex. names, SSNs, contact info, medical data). Do not share the underlying data via this form. C2C will reach out with secure data sharing methods.
5. What set of variables comprises the primary key that uniquely identifies a record in your dataset?
6. Does your dataset catalog information over time (time-series) or at a specific instance in time (cross-sectional)? Does your data cover multiple groups over multiple time periods (panel data) catalog information on one subject or multiple subjects? (ex. one school district vs. many school districts)
7. Does your dataset contain spatial data?

8. If "Yes" to the above questions, please provide information on the following:
- a. At what level is the full universe of your spatial data?
 - i. National/regional
 - ii. California
 - iii. California subdivision (ex. specific counties, regions within CA such as the Bay Area)
 - b. What is the primary geographic division within your data (ex. state, county, city, legislative district, LEA catchment area)?
 - c. Are there any secondary geographic divisions covered by this data? If so, please list them.
 - d. Is your spatial data in vector form (i.e. divisions are created by a series of interconnected points) or raster form (i.e. data is represented by discrete locations on a coordinate plane)?

Matching Instructions

9. Which of C2C's Data Providers do you intend to include in the matched dataset?
10. Which of C2C's Data Providers would be relevant to match on?
11. Which of the following matching frameworks best reflect your desired dataset?
 - a. One to one: each element in your dataset will match one element in C2C's dataset
 - b. Many to one: elements from one dataset will match to multiple elements in the other
 - c. Many to many: multiple elements from both datasets will match to multiple elements in the other dataset
12. Which of the following joins would you like us to conduct?
 - a. Outer join: records from at least one dataset will be fully preserved, even if they have no match
 - i. Which datasets should have their records fully preserved?
 1. C2C data
 2. Your data
 3. Both datasets
 - b. Inner join: only elements that are able to be matched will be preserved
 - c. Other join type (please specify)
13. Do you expect that the relevant data fields will match directly (ex. unique ID, SSN) or will the match be a fuzzy match (ex. names with different potential spellings)?
 - a. If you expect a fuzzy match, please provide instructions on how to adjudicate fuzzy matches (ex. prioritize one variable in the primary key, prioritize matches based on time, etc.)
14. Do you intend to match this data to any other publicly available data? If so, please provide information on this additional data.