Linking SSID and CCCApply Through Web Look-up Service

Business Rules for Technical Solution

**1.  Student submits an application for admission.**

a) Application is implied consent to permit CDE to share student data with CCCCO in order to evaluate the efficacy of educational programs.

**2. CCCApply calls a CDE web service (RESTful web service) passing (FirstName, LastName, BirthDate, HighSchoolCDScode, CCCID, SSID (optional).) *NOTE: Process uses exact match on all required fields.***

Input Definitions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Field | Description | Data Type | Length | Format |
| SSID | Statewide Student Identification Number | String | 10 |   |
| CCCID | California Community College Identification Number | String | 7 |   |
| FirstName | Student’s First Name | String | 30 |   |
| LastName | Student’s Last Name | String | 50 |   |
| BirthDate | Student’s Birthdate | String1 | 10 | yyyy-mm-dd |
| CDSCode | Student’s County-District-School Code | String | 14 |   |

a) SQL Server Date datatype, but for the purposes of an HTTP input, defined as string.

**3. CDE uses the data to match the student.**

    a) CDE stores the CCCID with their student record (if match is found).

    b) If match, CDE returns the SSID and CCCID to CCCApply via the web service. Else HTTP 400 response returned with “No student found” message.

Output Definitions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Data Field | Description | Data Type | Length | Format |
| SSID | Statewide Student Identification Number | String | 10 |   |
| CCCID | California Community College Identifier | String | 7 |   |

**4. CCCApply stores the SSID along with the CCCApply Application data as a field that is**

**downloadable by the colleges.**

**5. At a later date, CCCCO provides enrollment data to CDE and CDE can use CCCID and SSID for**

**matching.**

FAQs

1. QUESTION: With the attached business rules, what happens when the data sent by CCC Apply is not a perfect match?

ANSWER: In our initial discussions with Tim and his staff, we decided that we will do an exact match using the data provided. After we have gone through a cycle, we would take a look at the match rates and see if anything should be modified. FYI, I had one of my staff do a quick analysis of CALPADS data. We found that with the 12M+ student records that currently exist in CALPADS, 98.081% of the records have a unique combination of first name, last name and DOB. If you change the combination to first initial, last name and DOB, 93.789% of the records are unique. Of course, this does not take into account slight variations in spelling of names or misspellings, which leads us to the next question.

1. QUESTION: Does CDE have an algorithm to look at non-perfect matches that are highly likely to be the same student, or will this always be a perfect match or bust lookup?

ANSWER: As of now, we will only be looking at a perfect match. We haven’t had any discussion with Tim and his staff on looking for partial matches or variations in names. With our first cut, we are not using any built-in or third party phonetic algorithms (e.g. Soundex), but can implement it into the solution. With the initial discussions we had, we were going to go through an exact match process to see the accuracy and make adjustments. Making modifications to the matching process will be very straightforward.

1. QUESTION: If we use first initial, lastname, birthdate, and CDS code, would that get us a better match?

ANSWER: I don’t know if we would get a better match, especially if we get multiple matches. Below is the result for the first initial match. If we want to look at increasing the potential match rate, CALPADS has alias fields that can be used to match against. Even with exact matches, this should increase the percentage. We should probably have a discussion on the details of the match and what we changes we want to introduce. If we use Soundex (SQL Server), then we will be able to introduce phonetic matching to the names, which should increase the match rate. Keep in mind that as we get a little “fuzzier” on our criteria, we will be increasing the odds of getting false positives.

**First Initial + Last Name + Birthdate**

|  |  |  |  |
| --- | --- | --- | --- |
| **Result Count** | **Name Count** | **% of total** | **Running %** |
| 1 | 11080627 | 93.789% | 93.789% |
| 2 | 603303 | 5.106% | 98.896% |
| 3 | 92121 | 0.780% | 99.675% |
| 4 | 25139 | 0.213% | 99.888% |
| 5 | 8504 | 0.072% | 99.960% |
| 6 | 3072 | 0.026% | 99.986% |
| 7 | 1112 | 0.009% | 99.995% |
| 8 | 363 | 0.003% | 99.999% |
| 9 | 119 | 0.001% | 100.000% |
| 10 | 44 | 0.000% | 100.000% |
| 11 | 6 | 0.000% | 100.000% |
| 12 | 5 | 0.000% | 100.000% |
| 13 | 1 | 0.000% | 100.000% |
| 14 | 1 | 0.000% | 100.000% |