



*“We Help
Put America
Through
School”*

Data Strategy Enterprise-Wide Common Student Identifier

Matching Algorithm Working Session

April 28, 2003

Appendix B



Agenda

- Meeting Objectives
- High Level Design Phase Overview
- Matching Algorithm Discussion
- Next Steps



Session Objectives

- Review CSID High Level Design Approach
- Re-familiarize group with Matching Algorithm concept
- Discuss Matching Algorithm Rules, based on existing NSLDS algorithm
- Create draft of the matching algorithm for review by additional system experts and management



CSID Overview

What have we done? *Current State Analysis Phase*

- ✓ Completed a Current State Analysis of the FSA identification processes

Where are we going? *High Level Design Phase*

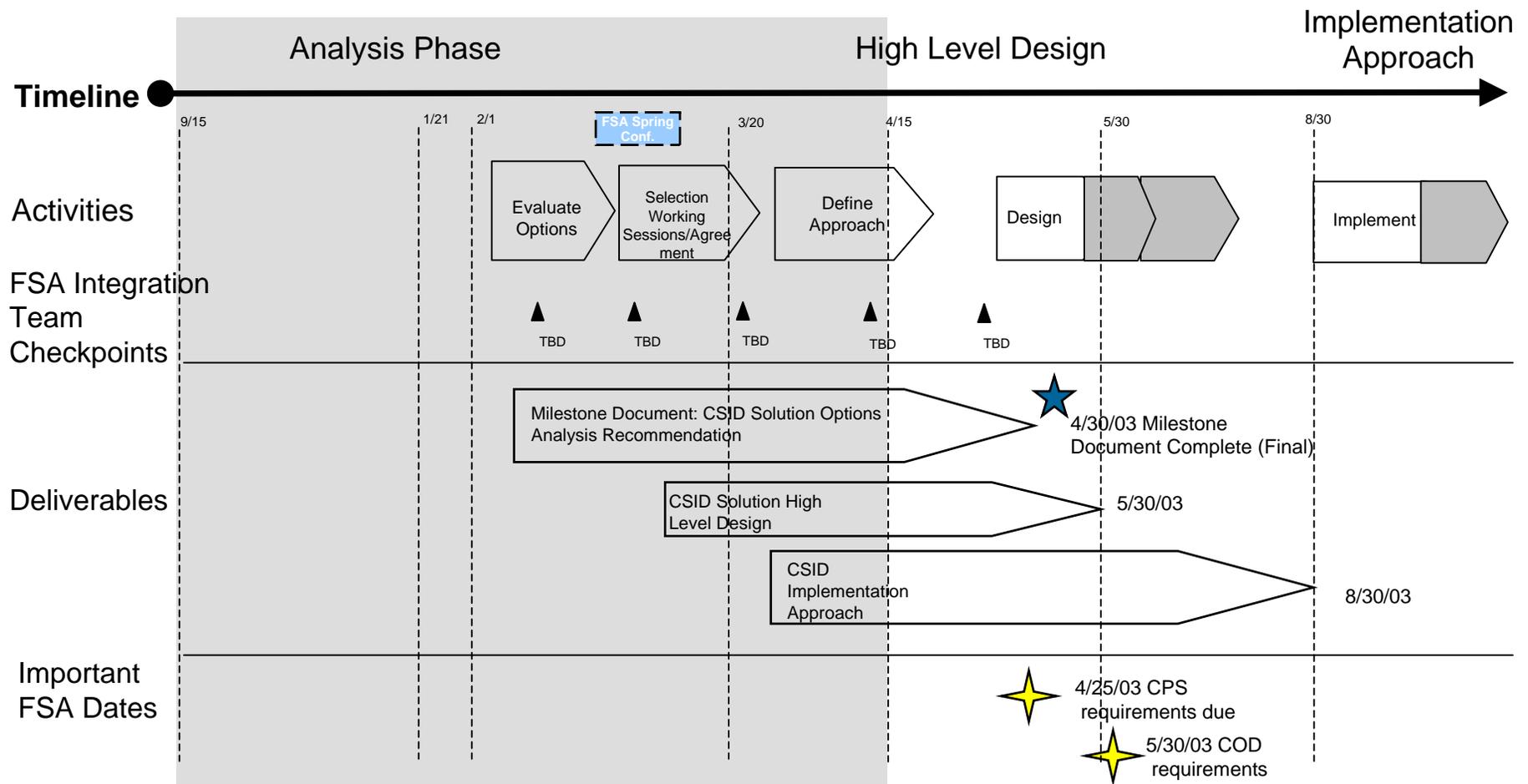
- ✓ Evaluate the CSID options
- ✓ Recommend a Common Student Identifier
 - Develop High Level Design and Implementation Approach for selected option

How are we going to get there?

- ✓ Facilitated working sessions with representatives from affected business owners within each major processing area (Aid Application, Common Services for Borrowers – [DLSS, DLCS, DMCS], COD, NSLDS) to analyze/review the potential CSID solutions.
- ✓ Conduct Consensus Meeting to review analysis and recommendations to decide on the preferred CSID solution.
 - Conduct Business Rules and Solution Design working sessions as needed.
 - Develop a High Level Design based on selected CSID solution.
 - Using the CSID High Level Design, the team will develop an Implementation Approach that supports the needs of the different FSA systems and business cycles.



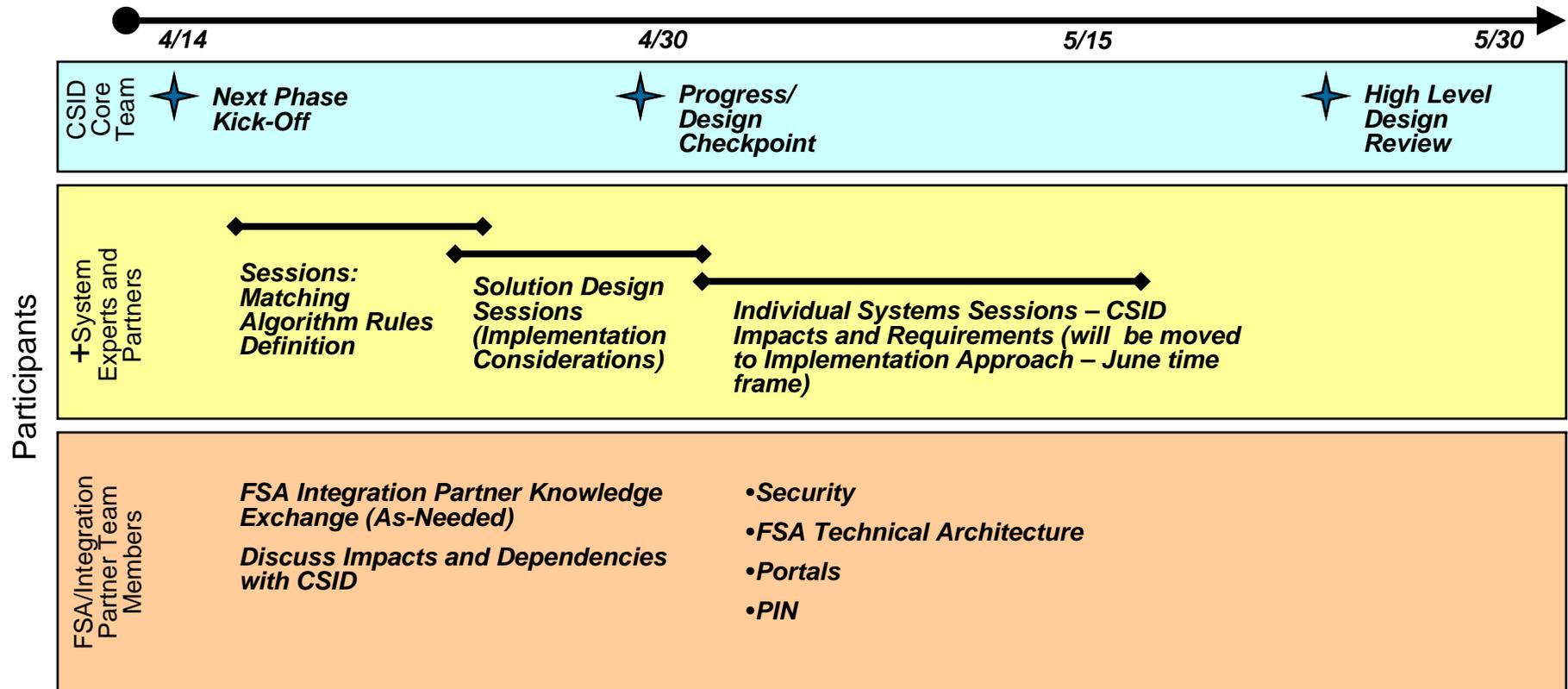
CSID Approach for High Level Design Phase





High Level Design Phase in Detail

Timeline





CSID Meeting Schedule - DRAFT

Meeting Topic/Major Activities	Recommended Attendees	Target Meeting Date
<u>CSID Team Introductory Meeting</u> : Familiarize CSID team members with the CSID High Level Design Phase going forward – particularly the definition of the Matching Algorithm	All CSID Core Team Participants	April 14
<u>Matching Algorithm Detailed Definition</u> : session(s) with groups of system experts and representatives	FSA System SMEs and CSID Core Team Members	Scheduled: April 28
<u>Solution Definition Session</u> : Core team session to discuss important implementation concerns: interfaces, testing, exception processing, etc.	All CSID Core Team Participants	Scheduled: May 5
<u>Individual System Requirements</u> : Assist each system with the detailed, specialized needs to implement the CSID solution.	FSA System SMEs and CSID Core Team Members	Moving to Implementation Approach - Summer 2003
Integration Partner/FSA Dependencies Checkpoints	Int Partner CSID Team and FSA Participants As-Needed	<i>As Needed</i> April 7 – May 30
High Level Design Document Review	All CSID Core Team Participants, BIG, Other Groups As-Needed	May 15-25



CSID Points of Agreement: 3/27/03

By employing a specific matching algorithm, or business rules, FSA systems can consistently identify customers using SSN and additional identifying information (DOB, First Name, and Last Name).

- Use of matching algorithm will be the most flexible way to compare and verify customer records before updates are made.
- The primary student identifier is SSN using a Matching Algorithm to provide additional verification checks on DOB, First Name, and Last Name.
- This solution option can be implemented in a variety of ways.
- Specific business rules will be defined in detail in the High Level Design phase.



Suggested Discussion Format

- Our starting point: Matching Algorithm rules will be based on the proven NSLDS algorithm
- Review existing NSLDS rules (Virginia scenario)
- Create Revised Rule (if necessary) pertaining to each data element
- Decide on the ordering and sequence of the rules



Discussion Summary

- **The following rules reflect the outcomes of the discussion.**
- **Many of the standards and rules used by NSLDS have been selected for the CSID algorithm.**
- **The matching algorithm will be a series of 4 comparisons of identifying data. Any one successful comparison constitutes a successful match.**

Additional Notes on the Rules:

- Matches on Names will not be case-sensitive.
- Fields like SSN may have consistent validity checks or 'standard edits' in addition to matching.
- Acceptable Plug Dates for the DOB match will be determined and discussed further (based on current processes and business needs of different systems).

First Comparison: SSN, First Name, and DOB



Data Element to Match	Rule
<p style="text-align: center;">SSN</p> <p style="text-align: center;"><i>AND</i></p>	<p>Current SSNs must match exactly on all 9 digits of the SSN on the student record.</p>
<p style="text-align: center;">First Name</p> <p style="text-align: center;"><i>AND</i></p>	<p>3 of the first 4 significant characters of the first name must match in sequence (in current or history), or alias matches exactly. Names of 3 characters or less must match exactly.</p> <p>Note: The phrase "in sequence" indicates that the letters must match in the same sequence. For example, Nary and Mary would match, as "ary" is in same sequential order. So would Maty and Mary, as "may" is in the same sequential order.</p>
<p style="text-align: center;">Date of Birth</p>	<p>Year matches exactly; or Year matches plus or minus one, with month matching exactly; or Year matches plus or minus ten, with month and day matching exactly; or Date is an acceptable plug date</p>

If the comparison of any of these rules is unsuccessful, the next set of rules are used for comparison.

Second Comparison: First and Last Names Transposed



Data Element to Match	Rule
<p style="text-align: center;">SSN</p> <p style="text-align: center;"><i>AND</i></p>	<p>Current SSNs must match exactly on all 9 digits of the SSN on the student record.</p>
<p>Incoming Record: Last Name Receipt Record: First Name</p> <p style="text-align: center;"><i>AND</i></p>	<p>Three of the first four significant characters of <i>last name on incoming record</i> must match in sequence (in current or history), the first name on the receiving record. <i>or</i> alias matches exactly.</p> <p>Names of 3 characters or less must match exactly.</p>
<p style="text-align: center;">Date of Birth</p>	<p>Year matches exactly; <i>or</i> Year matches plus or minus one, with month matching exactly; <i>or</i> Year matches plus or minus ten, with month and day matching exactly; <i>or</i> Date is an acceptable plug date</p>

If the comparison of any of these rules is unsuccessful, the next set of rules are used for comparison.

Third Comparison: First Initial Provided for First Name



Data Element to Match	Rule
SSN <i>AND</i>	Current SSNs must match exactly on all 9 digits of the SSN on the student record.
First Name <i>AND</i>	First name begins with same letter as first initial (a name that is an initial only or an initial followed by a period, not a comma).
Date of Birth	Day, Month, and Year match exactly.

If the comparison of any of these rules is unsuccessful, the next set of rules are used for comparison.

Fourth Comparison: First Initial and Last Name



Data Element to Match	Rule
SSN Match <i>AND</i>	Current SSNs must match exactly on all 9 digits of the SSN on the student record.
First Name Match <i>AND</i>	First character of first name matches first character of first name or first initial (current or history).
Last Name Match <i>AND</i>	Five of first seven characters of last name (excluding punctuation and spaces) match five of first seven characters of last name (excluding punctuation and spaces) (current or history). If fewer than five characters, all characters must match.
Date of Birth Match	Year matches exactly; <i>or</i> Year matches plus or minus one, with month matching exactly; <i>or</i> Year matches plus or minus ten, with month and day matching exactly; <i>or</i> Date is an acceptable plug date.

If the comparison of any of these rules is unsuccessful, exception processing must examine the records to resolve.

Outstanding Items for Solution Discussion: May 5, 2003



- Testing of the rules selected
- Use of Pseudo-SSNs and Pacific Islander identification numbers
- Use of Plug Dates for the Date of Birth match
- Corrections to SSNs for a student or customer
- Business processes that will require the matching algorithm to run
- Alias table – location, use, standardization, additions and updates for international names
- History of customer – locations of the histories compared during the matches
- Required field lengths for the algorithm
- Standard edits for validating fields: SSN, DOB, and Names
- Handling of customer records with only one name
- Exception and Error Processes for unsuccessful matches
- Reporting of matches that pass algorithm successfully, but may have small discrepancies (e.g. DOB typo - off by one year)



Next Steps

- Distribution of algorithm by CSID Core Team to their system experts
- Core Team collects comments – Please submit by May 7 –
- Team revises business rules to incorporate feedback
- Invitations will be sent for additional session to discuss Solution Design Issues scheduled for May 5 (9 am -1 pm)



Reference Slides

Discussion Notes – 4/28/03

Sample Matching Algorithm Flow

NSLDS Current Matching Algorithm

SSA Match Criteria



Suggested Rules: SSN

Example:	<i>SSNs must match exactly on all 9 digits of the SSN on the student record.</i>
A	<i>Current SSNs must match exactly on all 9 digits of the SSN on the student record.</i>
B	<i>Consistent checking a valid field range of SSNs</i>
C	<i>Consistent use of Pseudo-SSNs and Pacific Islanders' identification</i>
D	
E	



Suggested Rules: First Name

Example:	<i>3 of the first 4 significant characters of the first name must match in sequence (in current or history), or alias matches exactly. Names of 3 characters or less must match exactly.</i>
A	<i>Null or blank name would be the equivalent of NFN (Abbreviation of 'No First Name')</i>
B	<i>Not a case sensitive check for any name</i>
C	
D	
E	



Suggested Rules: Date of Birth

Example:	<i>Year matches exactly; or Year matches plus or minus one, with month matching exactly; or Year matches plus or minus ten, with month and day matching exactly; or Incoming DOB is real and NSLDS' DOB is one of the following plug dates: 19000101, 18991231, 18581117, 19581117, 19040404, 19600101, or 19??1111, where ?? can be any year.</i>
A	How should plug dates work? When will plug dates be permitted/prohibited?
B	In some situations the DOB match will be more stringent, because the name matches are based on less information (e.g. First Initial for First Name).
C	
D	
E	



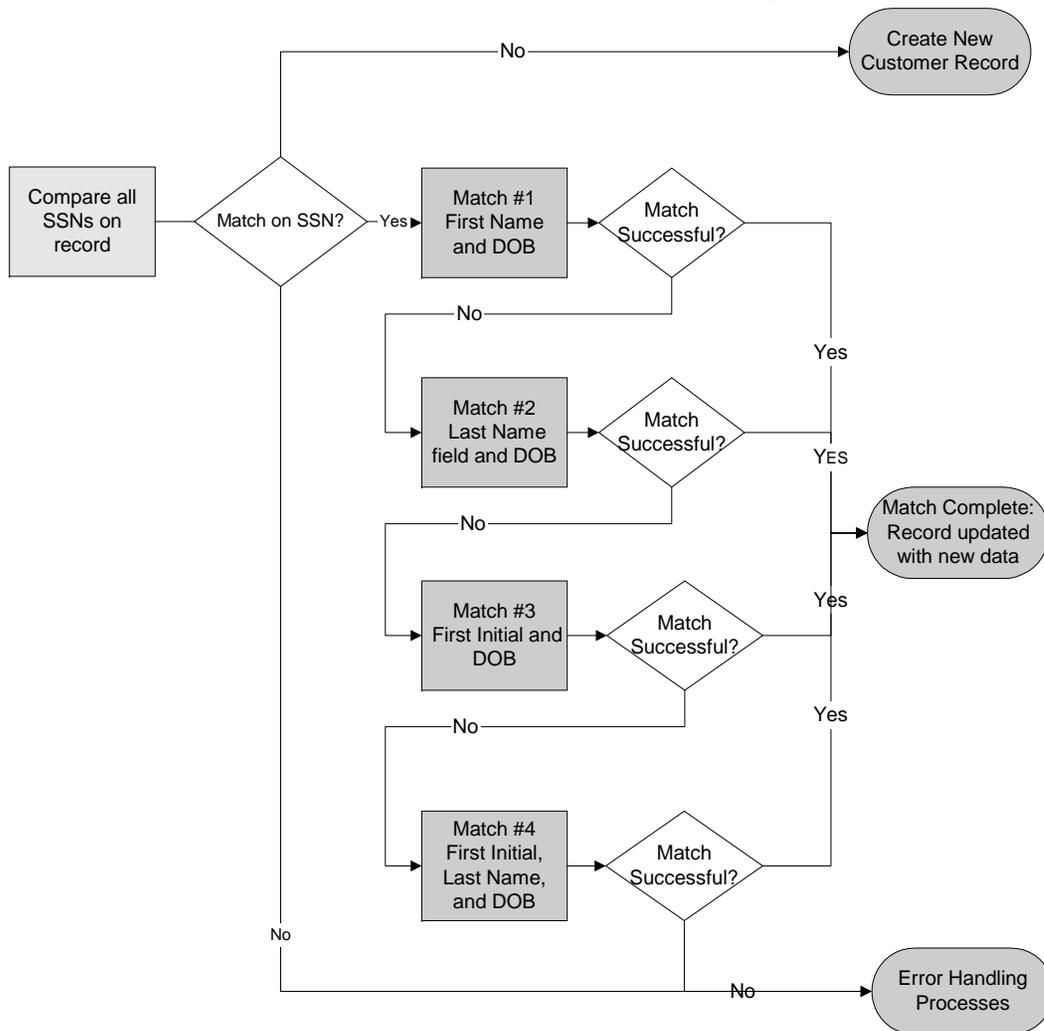
Suggested Rules: Last Name

Example 1: Used when first and last name may be transposed	<i>3 of the first 4 significant characters of the last name (in current or history) must match in sequence. Names of 3 characters or less must match exactly.</i>
Example 2: Used when first name match is only a single initial and DOB does not match	<i>5 of the first 7 significant characters of the last name (in current or history) must match in sequence. Names of 5 characters or less must match exactly.</i>
A	
B	
C	
D	



Matching Algorithm Sample

The following flow chart illustrates the way a sample matching algorithm might work, when one system receives a student record from another system:



Algorithmic Matches Explained:

#1 First Name and DOB Match

- First Name: first 4 characters, using current or history fields, and alias table
- DOB: real or plug date

#2 Transposed First Name & Last Name with DOB Match

- Last Name: first 4 characters, using current or history fields
- DOB: real, plug date, or some flexibility around year +/-1 and year +/- 10

#3 Match on First Initial of First Name and DOB (when only an initial and no other first names exists)

- First Initial: incoming first name begins with same letter as first initial using current or history
- DOB: an exact match and is not a plug date

#4 Match First Initial and Part of Last Name (with DOB Match)

- First Initial: first initial of first name matches first initial of first name or first initial using current or history
- Last Name: five of first seven characters of last name match five of first seven characters of last name using current or history
- DOB: real, plug date, or some flexibility around year +/- 1 and year +/-10



NSLDS Matching Algorithm (Reference)

First Name and Date of Birth (DOB) matches.

First Name

-Three of first four characters of first name on incoming record (excluding punctuation and spaces) match three of four characters of first name (excluding punctuation and spaces) in NSLDS (current or history), or alias matches exactly. The letters must match in the same sequence. If fewer than three characters, all characters must match; and

Date of Birth

Year matches exactly; or

-Year matches plus or minus one, with month matching exactly; or

-Year matches plus or minus ten, with month and day matching exactly; or

– Incoming DOB is real and NSLDS' DOB is one of the following plug dates: 19000101, 18991231, 18581117, 19581117, 19040404, 19600101, or 19??1111, where ?? can be any year.

Note: When NSLDS performs the analysis on the three of first four characters in first name or five of first seven characters in last name, the letters must match in the same sequence. For example, Nary and Mary would match, as "ary" is in same sequential order. So would Maty and Mary, as "may" is in the same sequential order

Transposed first name and last name with DOB match.

Last Name

-Three of the first four characters of last name on incoming record (excluding punctuation and spaces) match three of first four characters of first name (excluding punctuation and spaces) in NSLDS (current or history); and

Date of Birth

-Year matches exactly; or

-Year matches plus or minus one, with month matching exactly; or

-Year matches plus or minus ten, with month and day matching exactly; or

-Incoming DOB is real and NSLDS' DOB is one of the following plug dates: 19000101, 18991231, 18581117, 19581117, 19040404, 19600101, or 19??1111, where ?? can be any year.

Based on GA DPI document, section 6.5 Student Identifier, provided by NSLDS group



NSLDS Matching Algorithm (Reference)

Match on first initial of first name when NSLDS' first name is only an initial and no other first names exist in NSLDS.

First Name

Incoming first name begins with same letter as NSLDS' first initial (a name that is an initial only or an initial followed by a period, not a comma, and no first name in history); and

Date of Birth

Exact match and is not a plug date: 19000101, 18991231, 18581117, 19581117, 19040404, 19600101, or 19??1111, where ?? can be any year. (Note: If both incoming and NSLDS have same plug date, this is considered an exact match.)

Match on first initial and part of last name with DOB match.

First Name

First character of first name matches first character of first name or first initial (current or history); and

Last Name

Five of first seven characters of last name (excluding punctuation and spaces) match five of first seven characters of last name (excluding punctuation and spaces) in NSLDS (current or history). If fewer than five characters, all characters must match; and

Date of Birth

- Year matches exactly; or
- Year matches plus or minus one, with month matching exactly; or
- Year matches plus or minus ten, with month and day matching exactly; or
- Incoming DOB is real and NSLDS' DOB is one of the following plug dates: 19000101, 18991231, 18581117, 19581117, 19040404, 19600101, or 19??1111, where ?? can be any year.
- For loans or grants made before 1-1-1997, incoming DOB is plug date and NSLDS DOB is a real date.

Note: When NSLDS performs the analysis on the three of first four characters in first name or five of first seven characters in last name, the letters must match in the same sequence. For example, Nary and Mary would match, as "ary" is in same sequential order. So would Maty and Mary, as "may" is in the same sequential order.



SSA Match Criteria

- Goal is to verify that a given SSN is assigned to individual whose name is submitted.
- Verifies SSN based on:
 - Name tolerance rules
 - Exact agreement in the 1st seven positions of the last name and first two initials of first name; or
 - Exact agreement in the 1st seven positions of the last name and first initial of the first name matches SSA's first or middle initial, when only one initial is provided; or
 - Exact agreement in the 1st four positions of the first name or, if no first name is provided, first and middle initials match exactly and exact agreement in the 1st four positions of the last name; or
 - There is a one character difference or transposition of two adjacent letters in the last name and the first and middle initials match*; or
 - There is an extraneous character (I.e. JJOHNSO = JOHNSON) in the 1st seven positions of last name and first or middle initial matches; or
 - There is one missing letter in the 1st seven positions of last name and the first initial matches SSA's first or middle initial; or
 - Using some positional rules, a compound last name can be verified using only one of the surnames (I.e. Baker-Johnson, which only a match on Baker)
 - Date of Birth tolerance rules
 - Year matches exactly only (no match necessary on month or day); or
 - Year matches within range (+/-1) and month matches exactly
- Possible results include:
 - Verified
 - SSN not on file (I.e. never issued)
 - Name Matches, Date of Birth does not
 - Name does not match, Date of Birth not checked

*In this case, some additional leniency exists in first and middle initial matching.