



*“We Help  
Put America  
Through  
School”*

## **Data Strategy Enterprise-Wide Common Student Identifier**

## **Solution Design Working Session Outcomes**

May 5, 2003

May 15, 2003

Appendix C

# Agenda



Time	Topic	Location
5/5/03 9:00 - 9:30	Kick Off High Level Design Phase Overview	221 B & C
5/5/03 9:30 - 11:45	Solution Design Breakouts (and 15 min break) Group 1 – Reps from each system Group 2 – Reps from each system Group 3 – Reps from each system - Involves nearly all system handoffs/data exchanges - Approximately 10 people per group	Group 1 – 221 B Group 2 – 221 C Group 3 – 233
5/5/03 12:00 - 1:00	Consensus Meeting - Report breakout discussion - Discuss any discrepancies	221 B & C
5/15/03 1:00 – 3:00	Follow-Up Meeting -Discussion of changes and corrections -Discussion of exception processing	34D1

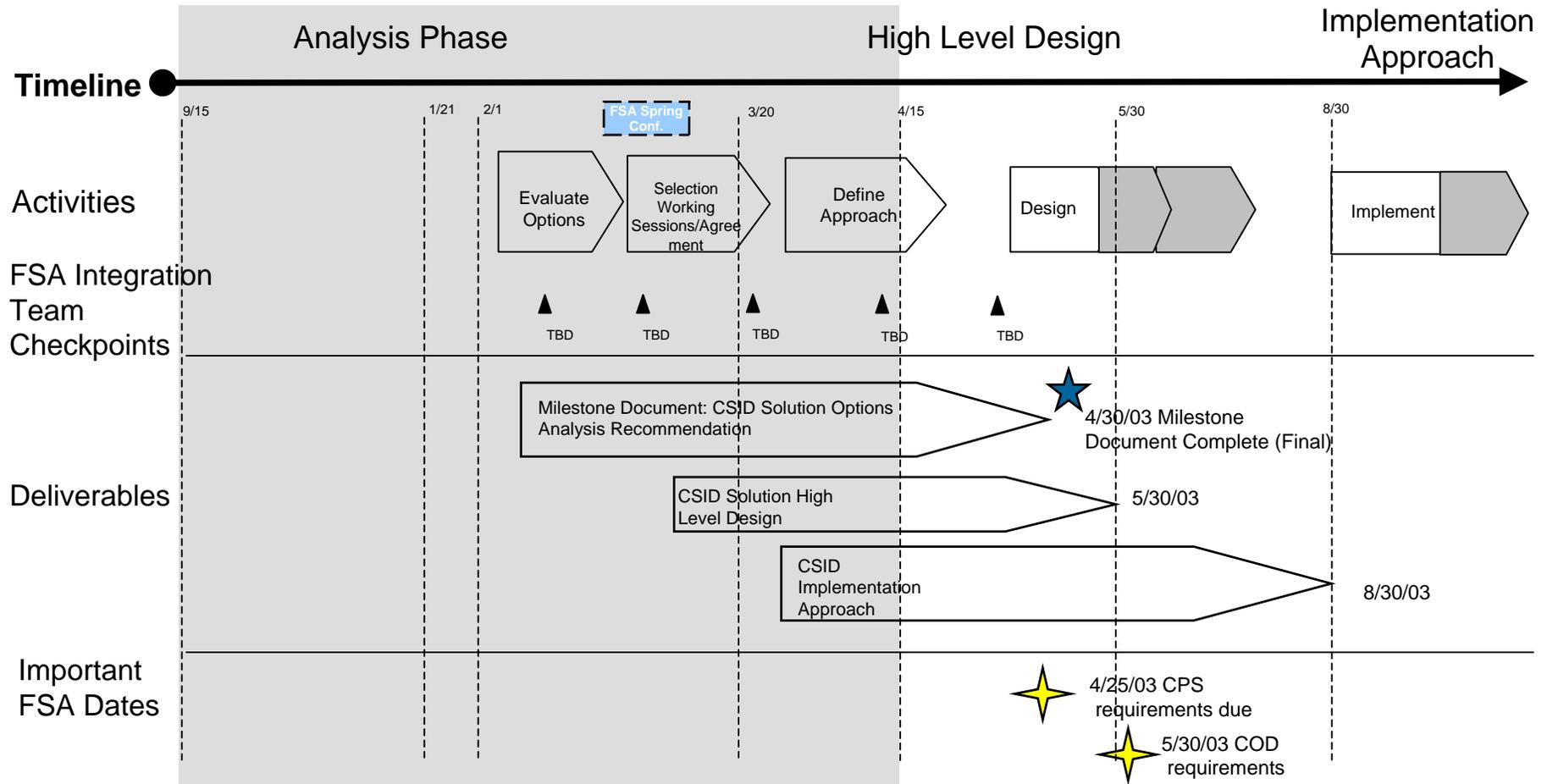


## Session Objectives

- Review CSID High Level Design Approach
- Re-familiarize group with key decision points for CSID
- Participate in Break-Out sessions where systems that must interface can all discuss issues related to CSID
- Create draft of resolutions for issues related to CSID
- Come to Consensus on preferred resolution(s)



# CSID Approach for High Level Design Phase



- Conduct Business Rules and Solution Design working sessions as needed. (4/27 and 5/5)
- Develop a High Level Design based on selected CSID solution. (Draft reviews TBD - 5/19-5/29)
- 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 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

Each breakout group should consider the processes and issues related to their specific topic.

- Try to agree on a recommendation/conclusion for all of the questions or issues related to that topic.
- Nominate a team recorder to maintain a written record of the feedback.
- Nominate a timekeeper to keep the discussion on track.
- If your group reaches an impasse, move on to the next topic/question in the list.
- There will be a short minute break at 11:45 to grab lunch.
- When considering the question, first examine the current method/process and modify it to suit the new solution.

*\*Note: The Matching Algorithm Rules were drafted in a separate meeting (held on 4/28/03). The outcomes of that meeting are included at the back of this document for reference; however, the specifics of the matching algorithm are NOT an agenda item for this particular meeting.*

# Topic One – Use of the matching standards



## High Level Topics

1. Instances in the lifecycle where the algorithm and SSA match be maintained or added to validate customer identity
2. Current interfaces and system controls that could enhance validation

Question	System to System	Centralized Source
<p><i>What points in the lifecycle should invoke the algorithm?</i></p> <p><i>What specific business processes/interfaces should invoke the algorithm?</i></p> <p><b>Note: NSLDS currently runs a matching algorithm whenever accepting or matching records with its system.</b></p>	<p><u>New Use of the Matching Algorithm:</u></p> <p>-All data exchanges for customer account information between CPS, PIN, COD, DLSS, DLCS, DMCS, NSLDS, CDDTS, SSCR.</p> <p>-Data received from external data providers (schools, FFEL community, etc.).</p> <p><b>Note: Systems that are part of CSB (CDDTS, DMCS, DLSS, DLCS) will eventually all be considered one single system.</b></p>	
<p><i>What specific business processes may not require the algorithm?</i></p>	<p>Current internal processes or processes between FSA systems that deal with subsequent payment posting or similarly specific transactions based on the exact matches of loan identifiers. [In these situations, the algorithm has been run when the new customer was added.]</p>	
<p><i>What additional points in the lifecycle would benefit from an SSA match?</i></p> <p><i>What specific business processes may require the additional SSA match?</i></p>	<p><u>Current Use of SSA Match Maintained:</u></p> <p>-FAFSA Applications to CPS</p> <p>-PIN Applications</p> <p><u>New SSA Match Added:</u></p> <p>-Direct PLUS Loans entering COD</p> <p>-Direct Loan Consolidation Application entering DLCS</p> <p>-FFEL PLUS Loans entering NSLDS</p>	
<p><i>When the algorithm is invoked, what history should be consulted (e.g. just the history in the receiving system, a centralized database of histories, etc.)?</i></p>	<p>For SSN and DOB, the matching algorithm should continue to compare the most recent (or current) SSN. The individual systems' histories of name should be consulted in the matching algorithm comparisons.</p>	<p>For SSN and DOB, the matching algorithm should continue to compare the most recent (or current) SSN.</p> <p>The cumulative history of name should be consulted in the matching algorithm comparisons.</p>
<p><i>When the algorithm is invoked, what alias table should be consulted (e.g. standardized tables in each system, centralized table, etc.)?</i></p>	<p>Content of the alias table should be managed and maintained by a single source and then distributed for loading into the individual systems as needed.</p>	<p>Content of the alias table should be managed and maintained at a single source that will load the most updated version to the centralized CSID solution.</p>
<p><i>What existing checks/interfaces could be enhanced to verify customer identification? (e.g. NSLDS – ISIR check)</i></p>	<p>NSLDS-CPS ISIR check could be used to flag identifier issues.</p> <p>NSLDS receipt of FFEL loans could check for a corresponding FAFSA record from CPS to verify SSA match flag OR add'l SSA Match may be needed.</p> <p>Matching algorithm CSID records could include the point in the algorithm where the match was successful to collect information about the different algorithm comparisons.</p>	

# Topic Two – Changes, Exceptions, and Errors



## High Level Topics

1. When the matching algorithm is unsuccessful, what should the overall approach be to handling the errors?
2. When changes or corrections are received, what process verifies the request as legitimate? Where should the changes or corrections be sent?

*The breakout group made the following high level decisions and agreed to resolve the more detailed questions during a follow-up discussion.*

- **Unsuccessful and successful partial matches will be involved in correction and exception processing.**
- **Consistent standards should exist regarding the supporting documentation required for changes to identifier information.**
- **Changes should be communicated to all phases of the lifecycle – a two way street – potentially communicated through the bus. All systems should be able to send and receive changes.**
- **Data should be cleaned and corrected in the front end of the lifecycle whenever possible to avoid bad data proceeding “downstream.” This may mean more errors in the beginning of the lifecycle: errors in front stage processing may result in rejection of data, and errors in late stage processing will result in cooperative error resolution**
- **COD Match Flag 4 discussion - COD is working with PDD to strengthen requirements around schools submitting borrower changes back to FSA when the borrower was not a Match Flag of “4.” One suggested strategy includes disbursing the loan for the initial occurrence, but sending schools a warning to correct the problem before the borrower’s next disbursement.**
- **External data providers will be impacted by changes and exception processing, and should be informed of the impacts.**
- **For each FSA system, dedicated resources must be identified to resolve errors and exceptions for the CSID.**
- **Resources should be devoted to a campaign that emphasizes and warns schools and the FFEL community about the gravity of CSID and the processing of good data, with valid SSNs, etc.**
- **FSA resources could communicate with the borrowers regarding the importance of submitting valid, correct data the first time, to avoid problems with processing aid.**

# Credentials and Documentation: Update to SSN



**What should the proof or standard be for accepting SSN changes? Sufficient support would be provided if**

The customer submits a current Social Security Card or Driver's License that displays the SSN

The customer's record completes a successful match of "4" with the SSA

Changes to SSN accompanied with changes to Name or DOB should be communicated in their entirety

System Receiving Change Request	Systems that must receive the updates					
	CPS	PIN	COD	CSB DLSS, DLCS, DMCS	NSLDS	OCTS 2.0
CPS		Updates currently received	Changes communicated with the AAR	Yes	No – new records are sent in the usual processing	No
PIN*						No
COD	Yes		Changes also sent to Schools	Yes	Yes, updates should be sent for Pell data	No
CSB – DLSS,DLCS,DMCS	Yes		Yes	Changes also sent to Credit Bureaus	Yes	No
NSLDS	No	No	No	No		No
OCTS 2.0	Return changes to source system of error for correction					

\*Due to the PIN re-engineering effort, the needs for changes and updates have not yet been incorporated; however, appropriate communications about changes and updates related to CSID will be established.

# Credentials and Documentation: Update to First Name



**What should the proof or standard be for accepting FN changes?**

Changes should be purposeful and customer-initiated. Changes to name should be sent forward in the lifecycle.

System Receiving Change Request	Systems that must receive the updates					
	CPS	PIN	COD	CSB DLSS, DLCS, DMCS	NSLDS	OCTS 2.0
CPS			Yes	Yes	Yes	No
PIN*						No
COD	Yes			Yes	No	No
CSB – DLSS,DLCS,DMCS	No		No		Yes	No
NSLDS	No	No	No	No		No
OCTS 2.0	Return changes to source system of error for correction					

# Credentials and Documentation: Update to Last Name



## What should the proof or standard be for accepting LN changes?

Legal documentation of the change, such as:

Marriage certificate, divorce decree, legal name change document

Receipt of change from a data provider requiring such documentation (e.g. FFEL community)

System Receiving Change Request	Systems that must receive the updates					
	CPS	PIN	COD	CSB DLSS, DLCS, DMCS	NSLDS	OCTS 2.0
CPS			Yes	Yes	Yes	No
PIN*						No
COD	Yes			Yes	No	No
CSB – DLSS,DLCS,DMCS	No		No		Yes	No
NSLDS	No	No	No	No		No
OCTS 2.0	Return changes to source system of error for correction					

# Credentials and Documentation: Update to DOB



What should the proof or standard be for accepting DOB changes?

Legal documentation of correction, such as:

Birth certificate

System Receiving Change Request	Systems that must receive the updates					
	CPS	PIN	COD	CSB DLSS, DLCS, DMCS	NSLDS	OCTS 2.0
CPS			Yes	Yes	Yes	No
PIN*						No
COD	Yes			Yes	No	No
CSB – DLSS,DLCS,DMCS	No		No		Yes	No
NSLDS	No	No	No	No		No
OCTS 2.0	Return changes to source system of error for correction					



# Reporting and Error Resolution

**Unsuccessful and successful partial matches will be involved in correction and exception processing.**

- What match scenarios should require a response or message to an FSA data provider?

Matches that have completely failed the algorithm

Potentially, imperfect matches that pass the algorithm with discrepancies (could be an optional request by systems).

- What match scenarios should require a response or message to an external data provider?

Matches that have completely failed the algorithm

*It will be important to prototype or test the logic of the algorithm first, to anticipate the errors and rejections prior to implementation.*

**For each FSA system, dedicated resources must be identified to resolve errors and exceptions for the CSID.**

- Are there current groups/teams devoted to such problems now?

- What primary functions do they perform?

- How could these groups' experience translate into CSID error resolution?

# Topic Three – Standardized Data



	Question	Response/Resolution
<b>SSN</b>	<i>What will the standard edits be for the SSN (e.g. must be within range issued by SSA)</i>	If a record being passed reflects a “4” SSA match, then the data can be considered verified; therefore, it may be necessary to include the Match Flag Value, as well as the date of validation, in passing data for the matching algorithm. If a record does not have a Match Flag of “4” on the record, the SSN should be checked against a valid field range.
	<i>How should the use of Pseudo-SSNs and Pacific Islander identification numbers be standardized? (e.g. pseudo SSNs from CPS, DMCS, NSLDS &amp; others?)</i>	Pacific Islanders receive different “888” SSNs every cycle year – these pseudo-SSNs should no longer be cycle year-specific.
	<i>Will current pseudos be acceptable when invoking the matching algorithm?</i>	Pseudo SSNs in the system can remain as is, but may require an indicator field that identifies the number as a pseudo, in the event that SSA begins to assign the current pseudo numbers.
	<i>Will future pseudos be acceptable when invoking the matching algorithm?</i>	Each system should be assigned a valid range of Pseudo SSNs. SSA has stated it will never assign SSNs beginning with “000.” This could become the standard prefix for pseudos.
<b>Name</b>	<i>How should the systems standardize handling of customer records with only one name?</i>	Systems should adopt the practice of populating the name fields in the way they are submitted by the customer. If the customer includes only a first or last name, the remaining empty field should be uniformly populated with “NFN” (No First Name) or “NLN” (No Last Name). The use of titles in name fields may need to be considered for Name as well.
<b>DOB</b>	<i>What will the standard edits be for the DOB (e.g. no customers older than 100 and younger than 13)</i>	The group suggested a range of 12-99 years of age as a valid age range, with the understanding that exceptions can be manually corrected for successful processing.
	<i>How should plug dates be standardized and recognized?</i>	Going forward, a single valid plug date, such as 19000101, should be decided upon throughout the enterprise, so plug dates are assigned uniformly.
	<i>When will “plug dates” be acceptable?</i>	Plug dates could be considered acceptable when coming from systems or partners who, for whatever reason, have lost or cannot provide the true DOB. Plug DOBs from the customer or borrower are not acceptable.
	<i>Will current plug dates be acceptable when invoking the matching algorithm?</i>	Plug dates that are currently used can be considered valid when checked against the algorithm for previous records.
	<i>Will future plug dates be acceptable when invoking the matching algorithm?</i>	The future records will use only a single valid plug date value OR may continue to use currently plug dates, but the use of existing plug dates would also require an indicator of the plug date as well as the source of the date.



## Next Steps

- Discussion of today's outcomes with additional system experts
- Core Team collects comments – Please submit by May 12 –
- Team revises to incorporate feedback into the High Level Design Deliverable
- Invitations will be sent to Core Team to review the progress of the High Level Design Deliverable, *tentatively scheduled for week of May 19*



## Matching Algorithm: 4/28/03 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).
- For each comparison, the match must be successful for all data elements compared.



# Matching Algorithm Comparisons

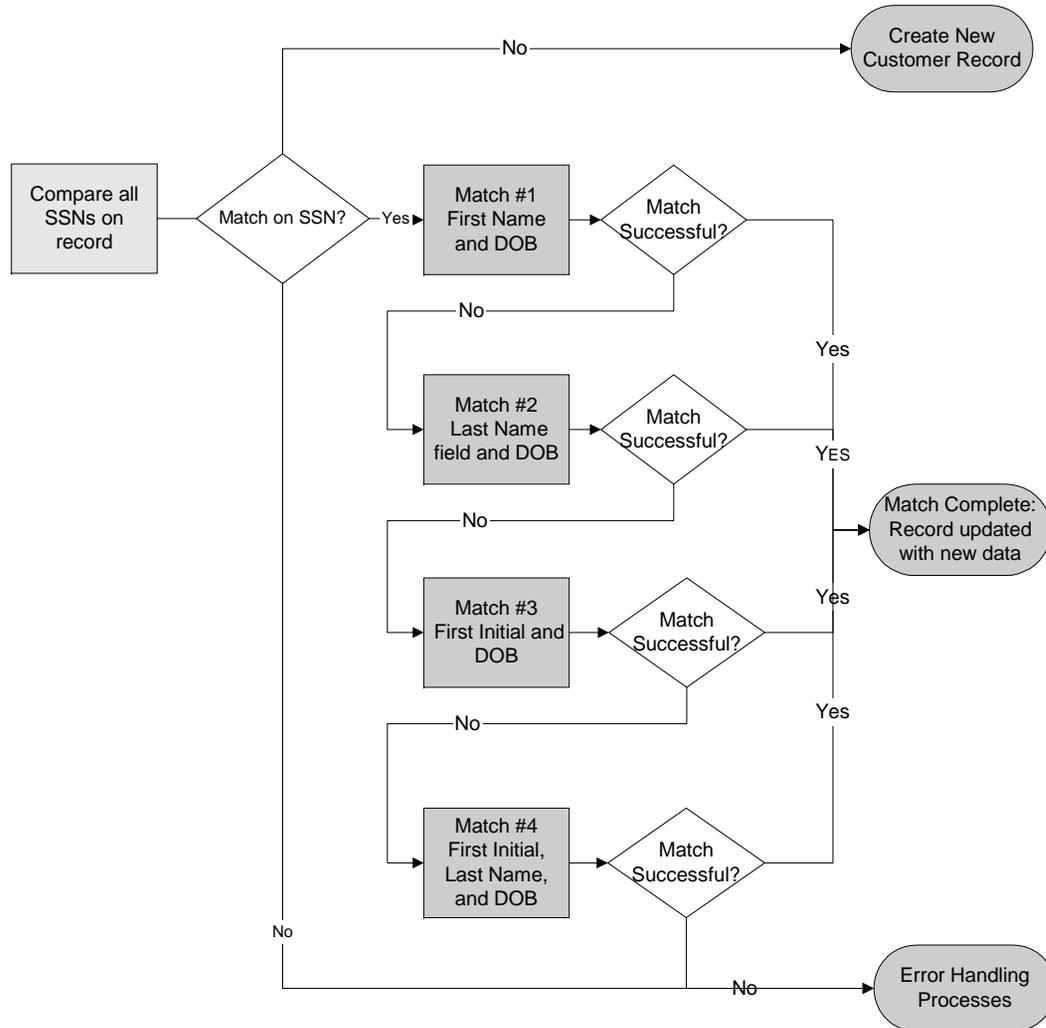
Comparison	SSN	First Name	Date of Birth	Last Name
<b>1<sup>st</sup></b> SSN, First Name, and DOB	Current SSNs must match exactly on all 9 digits of the SSN on the student record.	3 of the first 4 significant characters of the first name must match in sequence* (in current or history), <i>or</i> alias matches exactly. Names of 3 characters or less must match exactly.	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	<b>N/A</b>
<b>2<sup>nd</sup></b> Transposed First and Last Names	Current SSNs must match exactly on all 9 digits of the SSN on the student record.	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. Names of 3 characters or less must match exactly.	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	<b>N/A</b>
<b>3<sup>rd</sup></b> First Initial Provided for First Name w/ exact DOB	Current SSNs must match exactly on all 9 digits of the SSN on the student record.	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).	<b><i>Day, Month, and Year Match Exactly</i></b>	<b>N/A</b>
<b>4<sup>th</sup></b> First Initial Provided for First Name w/ check on Last Name	Current SSNs must match exactly on all 9 digits of the SSN on the student record.	First character of first name matches first character of first name or first initial (current or history).	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	Five of first seven significant characters of last name match in sequence (current or history). If fewer than five characters, all characters must match.

\*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.



# 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



# CSID 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 receiving system (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***



# CSID Matching Algorithm (Reference)

**Match on first initial of first name when receiving system's first name is only an initial and no other first names exist in receiving system.**

*First Name*

Incoming first name begins with same letter as receiving system's 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 receiving 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 receiving system (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 receiving system's 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 receiving 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 1<sup>st</sup> seven positions of the last name and first two initials of first name; or
    - Exact agreement in the 1<sup>st</sup> 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 1<sup>st</sup> four positions of the first name or, if no first name is provided, first and middle initials match exactly and exact agreement in the 1<sup>st</sup> 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 1<sup>st</sup> seven positions of last name and first or middle initial matches; or
    - There is one missing letter in the 1<sup>st</sup> 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.