



*“We Help
Put America
Through
School”*

Data Strategy Enterprise-Wide Common Student Identifier

CSID – Milestone Document

March 31, 2003

Appendix A



Milestone Document Contents

- Consensus Session Objectives
- CSID Overview
- Working Session Outcomes
 - Working Session Structure
 - Matching Algorithm (Business Rules) Overview
 - Matching Algorithm (Business Rules) Sample
 - Matching Algorithm (Business Rules) Scenarios
- CSID Points of Agreement
- Candidate Implementation Methods
- Implementation Method Feedback
- Implementation Considerations



Consensus Session Objectives

- Review Data Strategy and CSID background
- Review the outcome of the CSID Working Sessions – Matching Algorithm (Business Rules)
- Weigh Advantages and Disadvantages of the potential Implementation Methods
- Set direction on a preferred Implementation Method
- Discuss implementation considerations and potential process improvements to augment the CSID solution (e.g. error processing, enterprise change control)



CSID Overview

Objective The Common Student Identifier seeks to establish a simple framework by which FSA and Delivery Partners can consistently identify students/borrowers, across all phases of the Student Aid Lifecycle.

Key Problem Lack of enterprise-wide ID standards enables identification errors:

- Unique customer records can be inappropriately merged creating privacy concerns
- A customer's records can not be linking appropriately preventing FSA from viewing data about a customer across all phases of the life cycle

High Level Requirements

- Consistently and systematically link customer records across the FSA enterprise
- Support process changes and updates to key customer attributes (e.g. updates to First Name, Last Name, DOB)
- Ensure student privacy protection; minimize unauthorized/unauthenticated access to student data
- System identification requirements should not prevent valid customers from receiving aid or progressing through the repayment phase (e.g. deferments, rehabilitations, consolidations)



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*

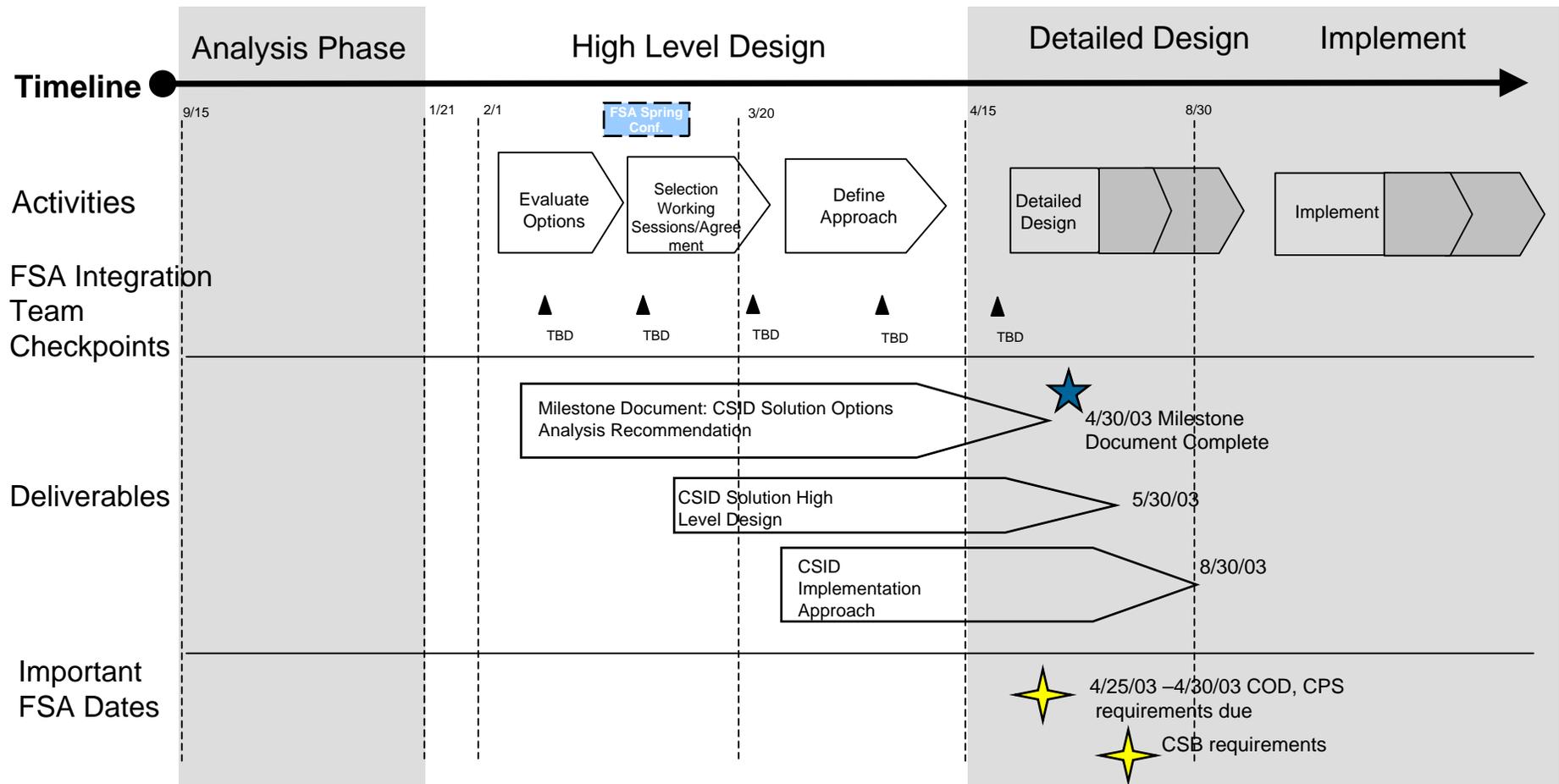
- ✓ Evaluate the CSID options
- ✓ Recommend a Common Student Identifier
 - Develop High Level Design and Implementation Approach for selected option

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.
 - 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 Overview





Working Session Outcomes

Working Session Structure

In each individual working session, the CSID Team used an Analysis Tool to discuss the CSID options with the Working Groups. The basic steps are as follows.

Evaluate the data elements

- Which pieces of customer data are most suitable as an identifier?

Discuss combinations of the data elements

- What data will be combined? How?

Determine the group's preferred option

- If we had to summarize the group's recommendation in 2-3 sentences, what would it be?

Discuss Implementation Methods

- What implementation method will be used?



Matching Algorithm Overview

The CSID Working Groups recommended using a Matching Algorithm (Business Rules).

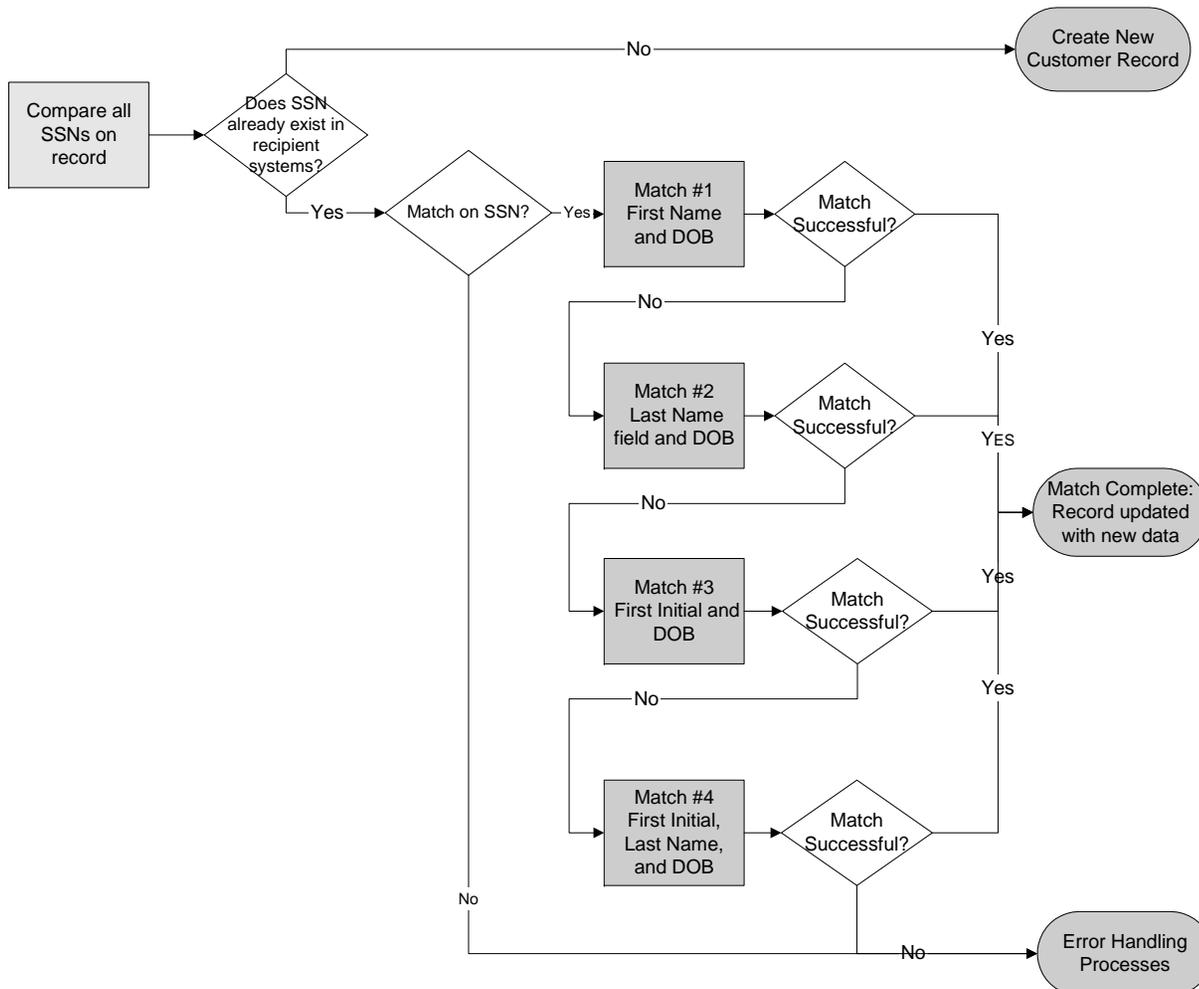
A static, physical identifier was discussed as a possible solution, but a logical identifier of existing fields was determined to be a more flexible option across the enterprise.

- Systems do not capture customer data or history consistently (i.e., DMCS and NSLDS may not have DOB).
- Therefore, physically using a combination of customer data elements is not an a viable enterprise solution.
- All of the Working Groups determined a Matching Algorithm was the most flexible CSID solution option for the FSA enterprise; however the implementation of that algorithm is varied.
- Recommendation: The primary student identifier is SSN using a Matching Algorithm to provide additional verification checks on DOB, First Name, and Last Name.
- For Today's Discussion: Feedback about the Matching Algorithm Concept, based on example that follows.
- *Note: Specific Business Rules will be defined in detail in the High Level Design phase.*



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



Matching Algorithm Scenarios

- 1. Problem: Inappropriately merging unique customer records resulting in privacy violation**
Customer 1 = 123456789 Anderson, Virginia 11/25/74
Customer 2 = 123456789 Anderson, Frank 7/20/50
Solution: The Matching Algorithm will reject or flag Customer 2 for error processing based on SSN, First Name and DOB match.

- 2. Problem: Not linking customer records correctly preventing FSA from viewing data across the loan life cycle**
Customer 1 = 999999999 White, Elizabeth 8/20/73
Customer 2 = 999999999 Burcaw, Beth 8/20/73 ('White' name change history captured)
Solution: The Matching Algorithm will marry the records based on SSN, First Name (alias table), Last Name (history change), and DOB match.



CSID Points of Agreement

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.



Implementation Method Breakout

Implementation Method	Description
Physical Use of CSID – <i>Not recommended</i>	<p>CSID will be the key identifier to be used internally within systems and externally to facilitate communication between systems.</p> <ul style="list-style-type: none"> •Working Groups did not recommend Physical Use of CSID several reasons: •FSA systems do not capture customer data and history consistently. •External Partners would not be able to use this option •Customers could experience more difficulty through the lifecycle, especially if ID data changes (e.g. Last Name)
Interface Support of Matching Algorithm	<p>CSID will be the key identifier to be used externally to facilitate communication between systems. External use of a single combination of identification data.</p> <ul style="list-style-type: none"> •Requires some changes to multiple FSA applications – mainly in their interfaces
Centralized Index for Matching Algorithm	<p>Creating an independent customer table to cross reference existing customers and history transactions.</p> <ul style="list-style-type: none"> •Requires an additional implementation of a centralized customer table •Enterprise program needed to manage index •Performance, data storage implications
Centralized Index and Row ID for Matching Algorithm	<p>Combination of creating an independent table to cross reference existing customers and the physical use of a CSID – Row ID.</p> <ul style="list-style-type: none"> •Requires an additional implementation of a centralized customer table •Requires each system to accept a new field – Row ID •Enterprise program needed to manage index •Performance, data storage implications



Implementation Method Breakout

The following key points were the result of the small groups' discussion about then potential CSID implementation methods.

	Interface	Centralized Index	Centralized Index/Row ID
Advantages	<p>Fewest changes to existing systems; easier to implement</p> <p>Most systems already capture all data required</p> <p>Exception processing maintained within each system</p>	<p>Single set of demographic history; single point of entry</p> <p>One owner of data; easier to diagnose problems and maintain data integrity</p> <p>Implementation of centralized SSA check to apply to all systems' data</p>	<p>Single set of demographic history; single point of entry</p> <p>One owner of data; easier to diagnose problems and maintain data integrity</p> <p>Implementation of centralized SSA check to apply to all systems' data</p>
Disadvantages	<p>No centralized tracking of changes or history</p> <p>Exception processing must be shared with other systems</p> <p>More difficult to have the same data across all systems</p> <p>Any changes to business rules must be updated across all systems' interfaces</p>	<p>Requires large modification/new system</p> <p>Potentially detrimental to performance</p> <p>Affects timeliness of error handling</p>	<p>Requires large modification/new system</p> <p>Potentially detrimental to performance</p> <p>Affects timeliness of error handling</p>
Impacts & Considerations Internal FSA Systems	<p>Implementation of standardized business rules for processing</p> <p>Use of data/keys on the web and PIN use</p> <p>Fewer changes/change requests between systems</p> <p>Must standardize the pseudo-ID process</p> <p>Difficult to keep systems' data aligned</p> <p>SSA Match Flag processing for updates</p> <p>NSLDS ISIR check on FFEL loans ?</p> <p>Coordinated prototyping/testing process</p> <p>Improves the situation, but will not solve all problems</p>	<p>Conversion of legacy data</p> <p>Implementation of standardized business rules for processing</p> <p>Not necessarily a new system</p> <p>Could be coordinated with Data Strategy on having a single owner for customer demographic data</p>	<p>Conversion of legacy data</p> <p>Implementation of standardized business rules for processing</p>



Implementation Method Breakout

	Interface	Centralized Index	Centralized Index/Row ID
Impacts & Considerations External Partners and Entities	<p>Harder to have integrated web access</p> <p>Might have to change school's acceptance criteria</p> <p>Implementation of standardized business rules for processing</p> <p>Impacts to current reporting practices and consolidation processing</p> <p>Long period of time for partners to fund and implement changes</p>	<p>Minimize/centralize authentication issues</p> <p>Cannot impact processing time for aid</p> <p>Implementation of standardized business rules for processing</p> <p>Counsels in favor of all loans going through the CPS/SSA match upon entry</p> <p>Long period of time for partners to fund and implement changes</p>	<p>Minimize/centralize authentication issues</p> <p>Cannot impact processing time for aid</p> <p>Implementation of standardized business rules for processing</p> <p>Counsels in favor of all loans going through the CPS/SSA match upon entry</p> <p>Long period of time for partners to fund and implement changes</p> <p>May have no impact to external agencies</p>
Cost/Benefit	<p>Least expensive option to implement, but maintenance could be add'l cost</p>	<p>More expensive – staff needed to maintain index</p> <p>All systems must develop interface to reach index data</p> <p>Do error rates warrant solution of this scale?</p>	<p>More expensive – staff needed to maintain index</p> <p>Implement additional field for Row Id to each system</p> <p>All systems must develop interface to reach index data</p> <p>Do error rates warrant solution of this scale?</p>



Implementation Considerations

Developing Enterprise Standards

Error Handling

- Enterprise error handling will be necessary.
- Standard processing may be necessary to consistently resolve identifier issues, based on current system exception processing.
- Business rules will be lifecycle dependent (i.e., defining an event that requires matching).

Change Controls

- Protocol for passing data and change history. (e.g. Would prior SSN be required as well as new SSN for a change request match?)
- Standards/rules for processing change requests by customers (e.g. name changes, SSN changes, DOB corrections).
- Use and assignment of pseudo or plug identifiers must be examined



Implementation Considerations

Process Impacts

- Additional SSA matches required at points in the lifecycle.
- Different processing of data may will be necessary for certain situations, dependent on FSA business needs.
- Algorithm may not be applied retro-actively
- CSID solution should not impede data flow

System Impacts

- CPS
- PIN
- COD
- DLSS
- DLCS
- DMCS
- NSLDS



Implementation Considerations

What must be considered/resolved outside the scope of this CSID solution?

- Authentication standards and consistency still required (PIN site)
- Treatment/cleansing of historical data



Sample 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.

Taken from GA DPI document, section 6.5 Student Identifier, provided by NSLDS group



Sample 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.