



**F E D E R A L
S T U D E N T A I D**

We Help Put America Through School

FSA Integration Partner

Data Strategy Enterprise-Wide

CSID

High Level Design

Deliverable 123.1.22

Version 2.0

June 16, 2003

Amendment History

DATE	SECTION/ PAGE	DESCRIPTION	REQUESTED BY	MADE BY
5/19/03	1	To shorten introduction - remove Data Strategy picture	R. Hartmuller	M. Picarello
5/19/03	2.1	Rephrase description of first bullet	R. Bowman	M. Picarello
5/19/03	2.1	Add explanation that CSID will not be applied retro-actively	P. Eliadis	M. Picarello
5/19/03	2.2	Age of NSLDS Matching algorithm changed from five to 10 years	M. Picarello	M. Picarello
5/19/03	2.2.1	Clarify 4 th comparison in Matching Algorithm Table	S. Martus	M. Picarello
5/19/03	2.2.2 and 2.3	Revise data flow pictures for readability	R. Bowman	M. Picarello
5/19/03	2.2.2	Bullet F-revise to reflect collections business need	S. Martus	M. Picarello
5/19/03	3.1	Move Implementation Options to Appendix	E. Dublin	M. Picarello
5/19/03	3.4	Add entry to Table 4-DLSS does not currently maintain 2 discreet fields for Last Name and First Name	R. Bowman	M. Picarello
5/19/03	3.6.1.2	Correct that Pacific Islander ID numbers can be re-used	D. Adams	M. Picarello
5/19/03	Appendices	Add mini-descriptions to meeting presentations	D. Adams	M. Picarello
5/19/03	Appendix E	Insert comment regarding real-time processing	D. Adams	M. Picarello
5/22/03	2.2	Insert explanations of figures for readability	T. Terry	M. Picarello
5/22/03	2.2, 3.3	General grammar and punctuation corrections	T. Terry	M. Picarello
6/4/03	2.2 Appendix F	Add information about risk of using more/less personal data elements for the identifier	D. Adams	M. Picarello
6/4/03	3.5.1.1	Correct typographical error	D. Adams	M. Picarello
6/4/03	3.5.1.1	Include match flag information on CSID record	D. Adams	M. Picarello

Table of Contents

1	PROJECT OVERVIEW	4
1.1	BUSINESS OBJECTIVE.....	5
1.2	CSID APPROACH AND PROGRESS	5
2	HIGH LEVEL DESIGN.....	7
2.1	CSID RECOMMENDATION	7
2.2	CSID MATCHING ALGORITHM	7
2.2.1	<i>Matching Algorithm Business Rules</i>	<i>8</i>
2.2.2	<i>Enterprise Use of the Matching Algorithm.....</i>	<i>9</i>
2.3	ADDITIONAL SSA VALIDATION	12
2.4	CSID SOLUTION DESIGN	13
2.4.1	<i>Uniform Demographic Change Process.....</i>	<i>13</i>
2.4.2	<i>Error and Exception Processing.....</i>	<i>14</i>
3	IMPLEMENTATION CONSIDERATIONS	15
3.1	PRIVACY, LEGAL, AND SECURITY.....	15
3.2	MAJOR POLICY IMPACTS.....	15
3.3	ADDITIONAL FSA SYSTEM IMPACTS	15
3.4	LINK TO AUTHENTICATION	16
3.5	STANDARDIZED DATA	17
3.5.1	<i>SSN</i>	<i>17</i>
3.5.2	<i>Names and Aliases.....</i>	<i>17</i>
3.5.3	<i>Date of Birth.....</i>	<i>18</i>
3.6	NEXT STEPS	18
	APPENDICES.....	19
A.	CONSENSUS MILESTONE DOCUMENT – MARCH 27, 2003	19
B.	MATCHING ALGORITHM SESSION OUTCOMES – APRIL 28, 2003	19
C.	SOLUTION DESIGN SESSION OUTCOMES – MAY 5, 2003 & MAY 15, 2003	19
D.	POTENTIAL IMPLEMENTATION OPTIONS.....	19
E.	LARGER VERSIONS OF CSID TARGET STATE DATA FLOW.....	19
F.	CSID PRIVACY ISSUE PAPER	19

1 Project Overview

Federal Student Aid (FSA) is seeking to deliver overall improvements in the areas of data quality and data consistency. FSA is focused on its overall approach towards data to ensure that accurate and consistent data is exchanged between its customers, partners, and compliance and oversight organizations. FSA will also leverage a targeted data strategy to support program-wide goals of maintaining a clean audit and removing FSA from the GAO high-risk list.

Senior FSA leadership has created a performance plan with several action items designed to take FSA off the GAO High-Risk List. The Data Strategy task order specifically addresses action item# 16. This action item identifies the need to define an enterprise-wide data strategy and high-level implementation approach that addresses the business flow of data across the enterprise, architecture, primary ownership, standards, management, access methods, and quality. The end result of the Data Strategy task order will be an overall enterprise data framework that integrates the following components that address FSA’s major data-related areas:

- Consistent Data Framework
- Technical Strategies
- XML Framework
- Common Identifiers
- Enrollment and Access Management

The Common Identifiers initiative is a key sub-item addressed by the FSA Data Strategy task order. The Common Identifiers initiatives are listed and mapped to the related deliverables within the Data Strategy work as illustrated in the following matrix:

Table 1. Deliverable Mapping

FSA ID No.	Action Item	Map	Del. Num	Name	Description
16.2.2	Develop requirements and initial design for Common Identifiers for School, Students.	→	123.1.22	CSID High Level Design	Documents the CSID High Level Design that will provide a consistent means of identifying students/borrowers across the Student Aid Lifecycle.
			123.1.25	RID Conceptual Design	Documents the RID Conceptual Design that will provide a consistent means of identifying entities across FSA's systems.
	Worked Performed by TO 123 in Addition To FSA's High Risk Plan		123.1.23	CSID Implementation Approach	Defines CSID sequencing and implementation strategy, dependencies, and key milestones for implementing the CSID Solution.

The Common Identifiers initiative includes the Common Student Identifier (CSID) and the Routing Identifier (RID). This deliverable focuses on the CSID High Level Design portion of the Common Identifiers initiative and seeks to provide a cross-system, common identifier strategy

for the student/borrower. The CSID High Level Design includes the CSID recommendation, matching algorithm business rules, and solution design considerations for implementation.

1.1 Business Objective

Currently, a lack of enterprise-wide ID standards prevents FSA from viewing data about a customer across all phases of the lifecycle. Lack of enterprise-wide ID standards generates identification errors:

- Unique customer records can be inappropriately merged creating privacy concerns.
- A customer's records cannot be linked accurately preventing FSA from viewing data about a customer across all phases of the lifecycle.

The CSID is a strategic component of the overall FSA Enterprise Architecture. The CSID initiative will help FSA view data across the student aid lifecycle, by isolating a primary identifier, accompanied by consistent business rules, which can be used by all FSA systems. The use of a uniform identifier solution will improve data integrity and reduce the current problem of merging and splitting student/borrower records due to identifier inconsistencies. In addition, a single CSID solution allows better analysis and reporting on the identifier problems, once discovered.

The CSID effort will:

- Define the high-level CSID design. This does not include a detailed-level design; rather, it will define available CSID implementation options and determine the functional requirements for the implementation of the CSID.
- Define the implementation strategy for the CSID that aligns with overall data strategy vision and business objectives.

The CSID initiative seeks to establish a simple framework by which FSA and Delivery Partners can consistently identify applicants and/or borrowers, across all phases of the student aid lifecycle. Such consistency will contribute to greater customer data quality and consistency.

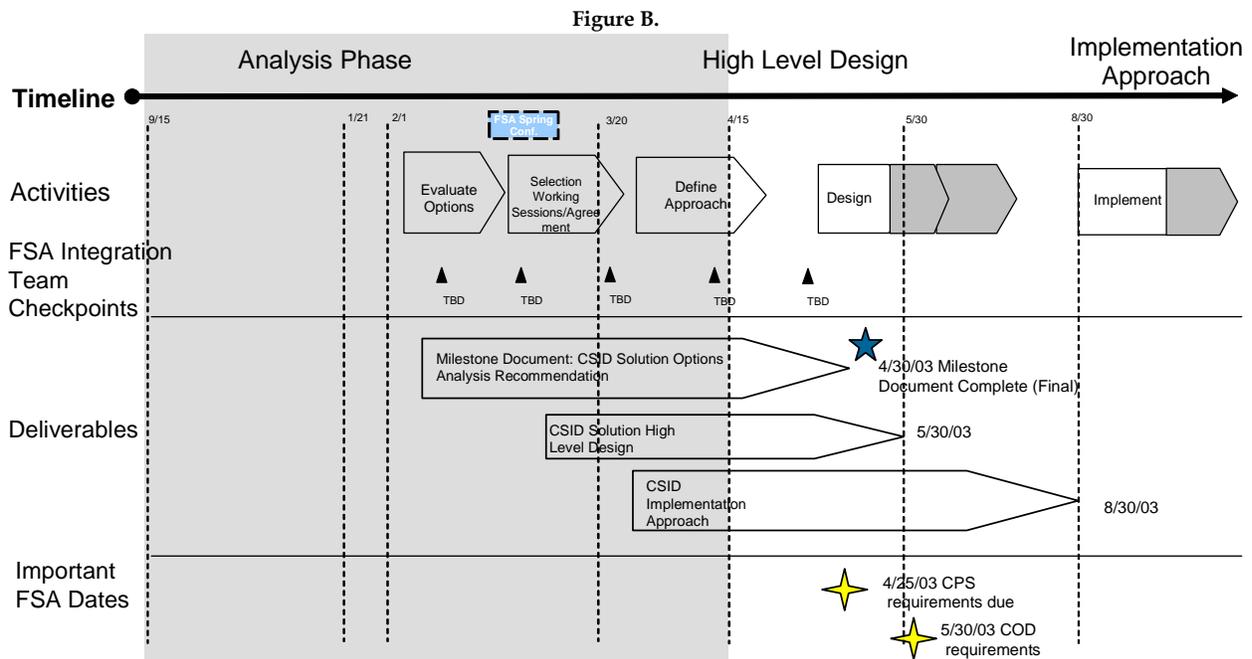
1.2 CSID Approach and Progress

The Common Student Identifier initiative is composed of multiple phases. The first phase consisted of a Current State Analysis (September 2002 - January 2003). During this phase, the CSID team documented the current methods and data elements used for customer identification with the major FSA systems. Specifically, the team met with representatives from Central Processing System (CPS), Common Origination and Disbursement (COD), Direct Loan Consolidation System (DLCS), Direct Loan Servicing System (DLSS), Debt Management and Collection System (DMCS), National Student Loan Data System (NSLDS), Ombudsman Case-Tracking System (OCTS), and PIN to gain an understanding of FSA's current student identification practices. The Current State Analysis was presented to the FSA business owners before proceeding with the current phase, High Level Design.

The CSID initiative formed a Core Team upon initiation of the High Level Design Phase in late January 2003. The purpose of the Core Team is to provide guidance and input to the CSID

initiative from each of the major FSA systems. The CSID Core Team plays an integral part in CSID progress. The Team includes at least one representative from each major system affected by the CSID (CPS, COD, DLSS, DLCS, DMCS, NSLDS, and PIN). The team has also included several representatives from Portals. Since inception, the Core Team has served as the primary point of contact for CSID direction-setting and decision-making. In addition, the Core Team reviews all CSID deliverables and work products before submission to FSA.

The High Level Design phase consisted of a series of collaborative working sessions in which the FSA system experts and business owners compared and recommended viable CSID solution options, determined matching algorithm business rules, and determined solution design considerations for implementation.



This document will summarize the CSID solution consensus reached over the course of several CSID working sessions around High Level Design with FSA business owners (*See Appendix A*). This document will outline the CSID Solution Design including the matching algorithm business rules and implementation considerations (*See Appendix B*). Additionally, the document will describe the process changes required to achieve a consistent identification solution across all the systems (*See Appendices C and E*).

Upon acceptance of this deliverable, the CSID Core Team will proceed with the Implementation Approach Phase. During this phase, the High Level Design will be used as a basis to draft a suggested approach for the implementation of the CSID solution. The Implementation Approach Deliverable will define CSID sequencing and implementation strategy, dependencies, and key milestones for implementing the CSID Solution in alignment with the overall data strategy and business objectives.

2 High Level Design

2.1 CSID Recommendation

After defining the current state in each system, the CSID Core Team examined the feedback expressed in each system's working session. The team considered a variety of viable options, including introducing a new identifier, using a combination of dynamic data fields for identification, and using a single data field for identification.

The team concluded that the preferred option needs to be both flexible enough to accommodate individual system's architecture, as well as tighten identification controls enterprise-wide. For performance and effort reasons, it was also concluded that the CSID solution is most effectively applied going forward, but not applied retro-actively to records already existing in FSA systems. The proposed solution that met these criteria is actually a three-pronged approach to the CSID solution.

1. The common identifier uses a combination of data fields common to all systems. The primary identifier is the Social Security Number, but it is verified through enterprise-wide business rules and tolerances with additional data, namely First Name, Date of Birth, and Last Name. (*Section 2.2*)
2. The common identifier will be verified with additional checks against the Social Security Administration where necessary. (*Section 2.3*)
3. The commonly experienced identifier corrections and changes will be resolved and communicated uniformly across all systems. (*Section 2.4*)

Throughout the High Level Design phase, the CSID Core Team has identified the key areas that will contribute to the High Level Design. The matching algorithm business rules, process changes, and system changes are described in the remaining sections of this deliverable.

2.2 CSID Matching Algorithm

The recommended CSID solution includes a combination of identifying data elements. The primary student identifier is Social Security Number (SSN), partnered with additional verification checks on Date of Birth (DOB), First Name, and Last Name using a matching algorithm with tolerances for common typographical errors. From a security perspective, the use of multiple pieces of personal data will be more difficult to defraud or falsify. This method is also commonly used in other financial service providers.

By employing a common set of business rules for identifying and linking student/borrower data, utilizing a shared matching algorithm, FSA systems can consistently identify customers using SSN and additional identifying information (DOB, First Name, and Last Name).

- Use of the 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.

For identification purposes, there are innumerable possibilities for matching algorithm rules. Financial institutions, credit bureaus, and federal agencies commonly use algorithmic business rules for identification verification. Within FSA, the National Student Loan Data System (NSLDS) employs a matching algorithm to verify borrower identities before updating system records. Over a period of nearly 10 years, the rules used by NSLDS have been revised and improved to reduce error rates while increasing successful identification verification. Consequently, the CSID Core Team began its matching algorithm discussion by fine tuning the proven rules used by NSLDS.

2.2.1 Matching Algorithm Business Rules

The CSID Core Team and system representatives met to compose the business rules for the matching algorithm; the following rules reflect the outcome of the discussion. Many of the standards and rules used by NSLDS have been selected for the CSID algorithm. A major change to the NSLDS algorithm is to only use the Current SSN field when verifying between systems. Currently, NSLDS also uses the SSN History field.

The matching algorithm will be a series of four comparisons of identifying data. Any one successful comparison constitutes a successful match, and for each comparison, the match must be successful for all data elements compared. These comparisons should preempt the majority of inappropriate merges of individuals' records, while preventing the fracture of a single person's information into multiple records. *For more details regarding these comparisons, see the CSID High Level Design Appendix B.*

In addition to these comparisons, the group discussed implementing overall data standards for the CSID fields (SSN, First Name, Last Name, and DOB). *For more details regarding these prospective data standards see section 3.6 of this document.*

Table 2. Matching Algorithm Details

Comparison	SSN	First Name	Date of Birth	Last Name
1 st 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), or alias matches exactly. Names of 3 characters or less must match exactly.	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	N/A
2 nd 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. or alias matches exactly. Names of 3 characters or less must match exactly.	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	N/A
3 rd 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).	<i>Day, Month, and Year Match Exactly</i>	N/A
4 th First Initial Provided for one of the First Names 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; 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	Five of first seven significant characters of last name match in sequence (current or history). If fewer than five characters, all characters must match.

2.2.2 Enterprise Use of the Matching Algorithm

Once the matching algorithm was determined, the CSID Core Team considered where the algorithm should be inserted into the FSA lifecycle. The Core Team also discussed the impacts of the CSID solution on FSA’s current identification processes. FSA system representatives contributed as the Core Team addressed these issues in the CSID Solution Design Working Sessions.

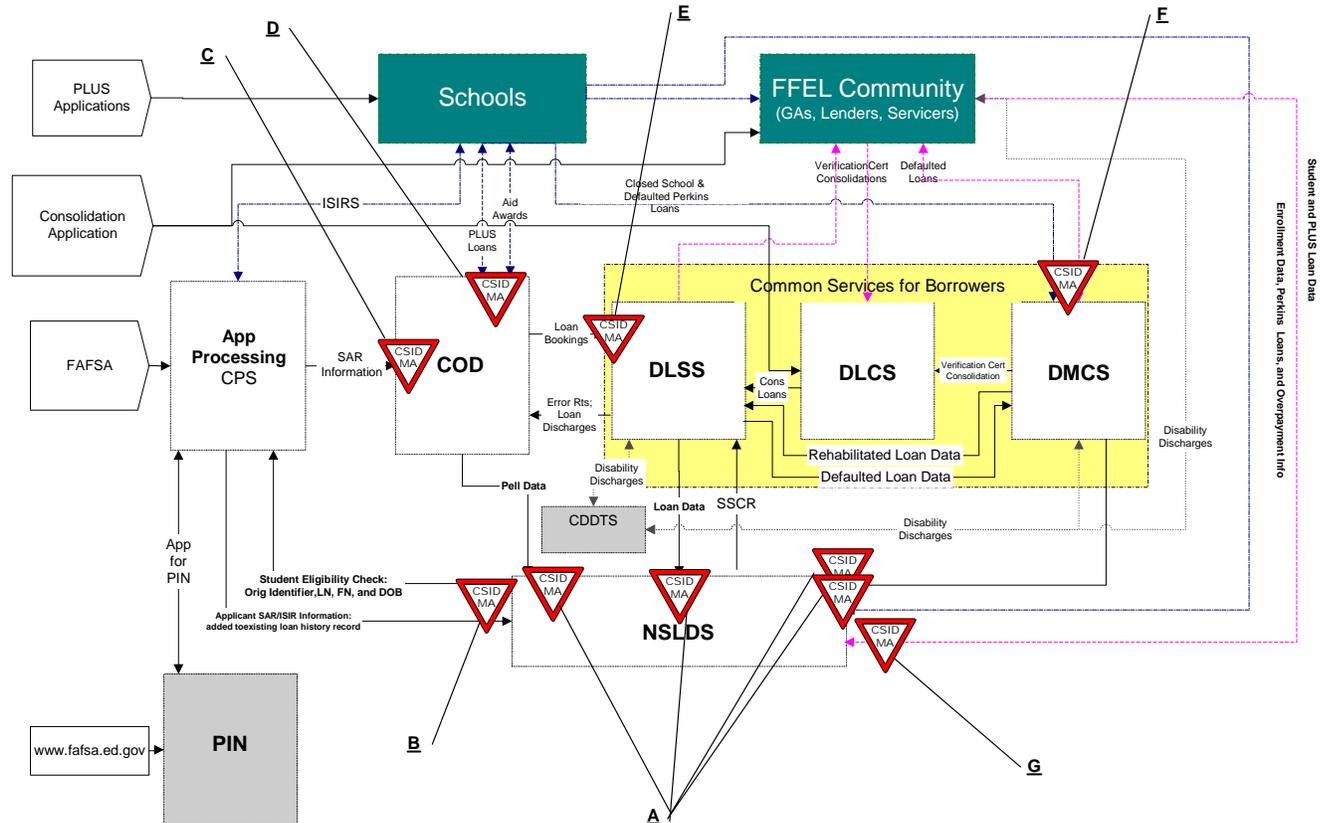
The CSID Core Team recommended several process changes affecting the way FSA systems communicate with each other and external entities regarding customer identifying information.

Regardless of the implementation methods, there will be changes and enhancements to the interfaces between the FSA systems as well as inputs from external partners. The High Level Design does not require external partners to make changes to the data they are sending or the format.

In the instance of interface support of the CSID, each system that collects or maintains student demographic data will be required to modify their interfaces with other systems, specifically the interfaces that enable the loading or updating of customer records. The interface modifications must include the enterprise business rules for the matching algorithm.

The following figure depicts the current state identifier data flow; interfaces are indicated by the lines entering and leaving system symbols. The matching algorithm should be inserted at the points indicated by the red triangles. Each letter indicates the point in the data flow that should add the matching algorithm to system processing.

Figure C. CSID Matching Algorithm and System Controls



Current:

A. NSLDS runs the matching algorithm for all new loan information entering NSLDS from the FSA systems and FFEL Community.

Proposed:

- B. CPS checks newly loaded FAFSA identity information against identification information in NSLDS; this interface already occurs, but could be modified to flag identification problems for exception processing.
- C. COD checks the AAR from CPS against existing COD records using the matching algorithm.
- D. COD runs the matching algorithm to verify the COD Aid Award records are updating the same identity (in addition to matching the Transaction number, as necessary). This data exchange may also provide opportunity to tighten the controls on the acceptance of records with an SSA match flag less than four (currently targeted for the 05/06 cycle year). (See Section 3.3)

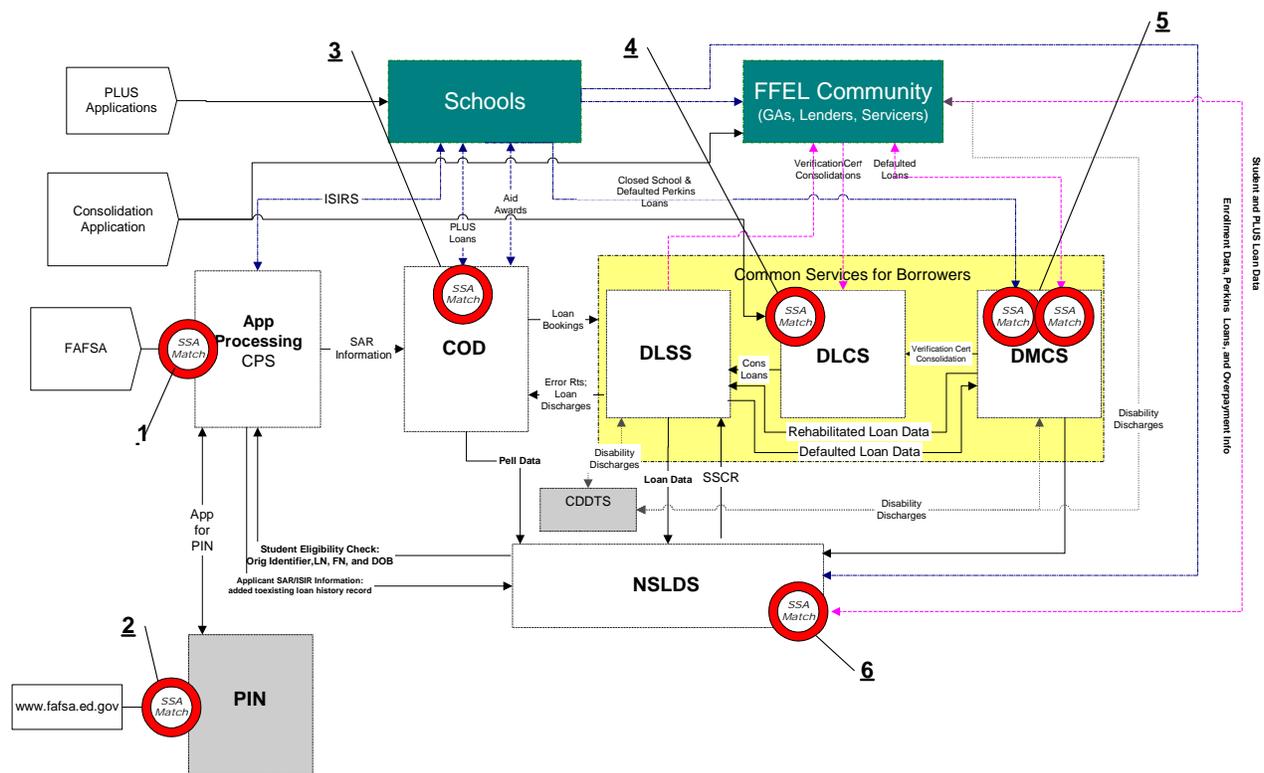
- E. DLSS (CSB) runs the matching algorithm to match records received from COD, DMCS, and DLCS with those existing in DLSS.
- F. DMCS (CSB) runs the matching algorithm to match debts received from FSA systems or the FFEL community with those existing in DMCS (in collections situations, debts with errors may be loaded to enable collection while completing exception processing).
- G. NSLDS verifies matching identifiers when matching ISIR records to loan records received from the FFEL community.

2.3 Additional SSA Validation

To ensure the creation of legitimate identities in the FSA systems, it is recommended that additional “SSA Match” interfaces be implemented. Depending on the implementation method chosen, the additional interfaces could be added in various places. In general, FSA should run a match at key points when customers may be entering the FSA systems with new or modified identifying information.

The following figure depicts the current state identifier data flow; interfaces are indicated by the lines entering and leaving system symbols. The SSA match should be inserted at the points indicated by the red circles. Each number indicates the point in the data flow that should add the SSA match to system processing.

Figure D. CSID Solution Design: SSA Matches



Current SSA Matches:

1. CPS Receipt of student FAFSA
2. PIN Receipt of application for the PIN

Proposed Matches:

3. COD Receipt of Direct PLUS aid award from schools
4. DLCS Receipt of the Consolidation Application
5. DMCS Receipt of new debt being loaded into the system
6. NSLDS Receipt of FFEL PLUS records from the FFEL community

2.4 CSID Solution Design

2.4.1 Uniform Demographic Change Process

Changes to a customer’s identifying information should be communicated to all necessary phases of the lifecycle; all systems should be able to send and receive such changes. It is not necessary to communicate every change to every system in every instance; however, such a capability must exist to ensure that important changes are communicated accurately.

Demographic changes to identifying information (SSN, Name, and Date of Birth) must be captured and communicated to all relevant systems at the time of receipt. The Core Team suggests requiring the following updates upon receipt of changes to SSN, Name, or DOB:

2.4.1.1 SSN Changes

The Core Team concluded that a change or correction to the Social Security Number is the essential demographic change requiring dissemination to all FSA systems. Consistent standards should exist regarding the supporting documentation required for changes to SSN information. The Core Team suggested that the following current verifications used by systems are acceptable:

1. Submission of a valid Social Security Card or Drivers License that displays the Social Security Number.
2. Receipt of a successful SSA match (match flag of 4).
3. Change request received from a data provider who requires similar credentials.

System receiving the initial request	Systems that must receive updates				
	CPS	PIN	COD	CSB (DLSS, DMCS, DLCS)	NSLDS
CPS ¹			X	X	X
PIN ²					
COD	X			X	
CSB (DLSS, DMCS, DLCS)	X		X		X
NSLDS					

¹ This provision assumes that CPS becomes a multi-year database, making their receipt of SSN changes applicable across multiple award years.

² 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.

2.4.1.2 Name and DOB Changes

The CSID Core Team concluded that Name and DOB changes should move forward through the lifecycle as they do today. Such changes may also be sent backward through the lifecycle if the system or business need requires it.

1. In the instance of a last name change, proof of a marriage license, divorce decree, or legal name change document.
2. Change request received from a data provider who requires similar credentials.
3. Dates of Birth corrections do not require additional documentation.

System Receiving the Initial Request	Systems that must receive updates				
	CPS	PIN	COD	CSB (DLSS, DMCS, DLCS)	NSLDS
CPS			X	X	X
PIN					
COD	X			X	
CSB (DLSS, DMCS, DLCS)					X
NSLDS					

To enable the most accurate change information, the communication of such identifier changes should include:

- Original or previous CSID data
- Corrected or revised CSID data
- Date/time the change was received
- Source of the change request

2.4.2 Error and Exception Processing

Detailed process and system requirements will be identified as the CSID initiative develops Implementation Options. The following high-level requirements were developed during the CSID Working Sessions:

- Unsuccessful matches for both the matching algorithm and the SSA Match will be included in correction and exception processing. Successful but partial matches may also be included in the exception processing.
- 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.
- External data providers may be impacted by changes and exception processing, and should be informed of the impacts. (e.g. The processing or turn-around time for error resolution may be different from a partner’s current process.)

- 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 cautions schools about the data integrity benefits of CSID and the processing of good data, with valid SSNs, etc.
- FSA resources should communicate with the borrowers regarding the importance of submitting valid, correct data the first time, to avoid problems with processing aid. (e.g. A message could be included on the paper and web applications for aid (FAFSA, PLUS Application, etc.).)

3 Implementation Considerations

3.1 Privacy, Legal, and Security

FSA Integration Partner is currently developing a Technical Security Framework. CSID's selection must be aligned with the Identity Access and Management portion of the CSID. The CSID team has met with the security team on several occasions and is in alignment to date. (See *Appendix F: CSID Privacy White Paper*)

3.2 Major Policy Impacts

- Integration with the reengineering of the PIN/online authentication
- Requirement and implementation of the SSA match at new points in the FSA Enterprise (see section 2.3.1)
- COD's desire to tighten the requirements for schools to correct and reconcile identification errors before being accepted from CPS (in discussions with PDD)

3.3 Additional FSA System Impacts

Some systems' data structures do not currently support the CSID solution, so the following functional requirements have been identified. These changes will be imperative prior to the implementation of the CSID.

Table 4. CSID System Requirements

System	Conflict with CSID	CSID Requirement
CPS	<p>Original ID on FAFSA cannot be changed:</p> <ul style="list-style-type: none"> ▪ Changes to name and SSN are not reflected in the Original ID. ▪ Any duplicates of Original ID are treated as subsequent applications (new transactions on existing applications). ▪ When an applicant uses the SSN of another applicant, the applicant in error must re-apply for aid to receive correct Original ID. <p>It is possible for Current SSN (NOT Original ID) to be duplicated in CPS. Applicant records in CPS are not linked across cycle years.</p>	Modify its system to be unique on the Current SSN.
COD	<p>Records with imperfect SSA matches are often sent from Schools to COD; Records that "fail" the SSA match are not sent to</p>	Develop a process to work and resolve records rejected from COD because of bad identifying information

System	Conflict with CSID	CSID Requirement
	COD; however, COD does receive records that receive a partial match at SSA.	(e.g. bad SSA match flag). 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. COD would send school updates to downstream FSA systems.
COD	Name changes received by COD are updating the appropriate record as a maintenance record, rather than a validated change.	Attribute multiple last names to a customer record instead of creating two unique customer records.
DLSS	First and Last Names are not maintained in discreet fields in the DLSS system. The system has a single Name field.	Modify/split field when communicating to other systems to distinguish First Name and Last Name.
PIN	SSN, DOB, and/or Last Name changes require re-application for a new PIN. The record's unique identifier (stable data) cannot be duplicated within the PIN database.	Maintain unique SSN in database (<i>for more details, see Section 3.5 of this document</i>).

The new release of the FSA Students Portal may collect the SSN and additional identifying information for students, once they enter the aid awareness process. This identifying data will be linked to the rest of the CSID once submitted through a FAFSA.

Prior to the submission of the SSN and other key identifying information, the Portals system will not require an interface or affiliation with student records held in the CSID solution systems. Once the key identifiers are given, students who have logged into the Portal can be associated with their records in CSID affected systems, for tracking and research purposes.

3.4 Link to Authentication

The PIN re-engineering process must be in synch with the decisions regarding the CSID. Current PIN functionality does not facilitate the future CSID; however both initiatives are tracking their respective interdependencies.

The PIN database is keyed on the combination of SSN, DOB, and the first two letters of last name; therefore, if any one of those three elements changes, a new record is created in the database. To adequately support the CSID, the PIN database must be unique on SSN. Additionally, PIN should allow changes and corrections to names and DOBs without creating a new PIN to improve customer service (e.g. customers will only require a single PIN to view data in the event of a name change).

3.5 Standardized Data

The CSID Working Session participants identified a need for specific data standards relating to the customer identifying fields (SSN, Last Name, First Name, and DOB). The standards for this data are one aspect of the FSA Data Strategy Enterprise-Wide initiative; therefore, the standards themselves will be defined outside this initiative. Suggested standards pertaining to the CSID data elements are explained below.

3.5.1 SSN

3.5.1.1 *Valid with SSA*

Records that contain a “good” SSA Match Flag (4), the SSN can be trusted as legitimate since it has already been verified with the Social Security Administration. If a record does not have a Match Flag (4) on the record, the SSN should be checked against a standard valid field range. Currently, different FSA systems maintain different valid field ranges; therefore, the valid range must be standardized. In order to rely on this verification, the match flag must be sent or associated with the customer record as it proceeds through the lifecycle.

3.5.1.2 *Pacific Islander*

Pacific Islanders presently receive “888” SSNs every cycle year; these Pseudo SSNs can be cycle-year-specific. Alternatively, if the applicant remembers his/her Pseudo SSN from the previous year, the Pseudo can be re-used. The existing Pseudo SSNs are easily identifiable, since the numbers are currently not issued by SSA; however, these records may require an indicator field that identifies the number as a pseudo, in the event that SSA begins to assign the current Pseudo SSNs to the general population.

Pacific Islanders should receive a single Pseudo SSN that will be maintained as their unique customer identifier throughout the customers’ lifetime interaction with FSA.

3.5.1.3 *Lost or Corrupted Customer Records*

Pseudo SSNs are often assigned when borrower records are incomplete, usually due to a collections or closed school situation. The existing Pseudo SSNs are easily identifiable, since the numbers are currently not issued by SSA; however, these records may require an indicator field that identifies the number as a pseudo, in the event that SSA begins to assign the current pseudo numbers.

In the future, each system should be assigned a valid range of Pseudo SSNs. SSA has stated it will never assign SSNs beginning with “000;” therefore, this could become the standard prefix for Pseudo SSNs.

3.5.2 Names and Aliases

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).

To prevent rejection and identification conflicts due to common nicknames and short-forms of names, the CSID matching algorithm will compare first names to a standardized Alias Table. 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.

3.5.3 Date of Birth

3.5.3.1 *Valid Range*

FSA systems maintain different valid ranges for the age of a borrower. The group suggested a standard range of 12-99 years of age, with the understanding that exceptions can be manually corrected for successful processing. COD suggested a lower age range, of eight years, due to the young age of some aid recipients.

3.5.3.2 *Plug Dates*

In the event that a borrower's record does not have Data of Birth, systems commonly populate the DOB field using a "Plug Date." The Plug Date is typically a combination of numbers that are easily identified (e.g. birth day begins with '88' or birth year is '1900').

Going forward, a single valid plug date, such as 19000101, should be decided upon throughout the enterprise, so plug dates are assigned uniformly. The DMCS would prefer a default value of 00000000 rather than a formatted "dummy" date. The enterprise consensus on these values will be discussed further in later phases of the project. 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 customers or borrowers are not acceptable.

Plug dates that are currently used can be considered valid when checked against the algorithm for previous records. 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.

3.6 Next Steps

From June to August 2003, the CSID team will focus on the Implementation Approach for the CSID solution. Several important options and points of discussion have emerged in the course on the High Level Design Phase (*See Appendix D*).

Appendices

- A. *Consensus Milestone Document - March 31, 2003***
Documents the discussion and outcomes from the CSID Consensus Meeting, where the preferred CSID solution was selected.
- B. *Matching Algorithm Session Outcomes - April 28, 2003***
Documents the discussion and outcomes from the CSID Core Team Meeting, where the matching algorithm business rules were defined.
- C. *Solution Design Session Outcomes - May 5, 2003 & May 15, 2003***
Documents the discussion and outcomes from the various solution design meetings, seeking to resolve the major impacts of the CSID on FSA systems and lifecycle.
- D. *Potential Implementation Options***
Outlines the potential CSID Implementation Options.
- E. *Larger Versions of CSID Target State Data Flow***
Gives enlarged image of Figures C and D of this deliverable.
- F. *CSID Privacy Issue Paper***
Documents the guidelines and laws that may impact the use of Social Security Numbers, or Tax Payer Identification Numbers, as part of a Common Student Identifier for FSA.