



F E D E R A L
S T U D E N T A I D
We Help Put America Through School

FSA Modernization Partner

NSLDS II Reengineering Data Conversion and Migration Strategy

Version 1.0

July 19, 2002

Table of Contents

1	INTRODUCTION	2
1.1	OBJECTIVE	2
1.2	BACKGROUND.....	2
1.3	SCOPE.....	2
1.4	CONVERSION ARCHITECTURE	3
2	DATA CONVERSION APPROACH	5
2.1	OVERVIEW.....	5
2.2	INITIAL DATA CONVERSION.....	5
2.2.1	<i>Data Extraction</i>	5
2.2.2	<i>Data Transformation</i>	8
2.2.3	<i>Data Load</i>	10
2.3	SECONDARY DATA CONVERSION.....	10
2.4	AGGREGATE TABLES.....	11
2.5	DATA VALIDATION.....	12
2.6	ARCHIVE STRATEGY/CONVERSION.....	13
2.7	ONLINE STATISTICAL ABSTRACT CONVERSION.....	13
2.8	SYSTEM TEST	13
2.9	ASSUMPTIONS	14
2.10	RISKS	15
3	DATA CONVERSION PLAN	15
3.1	OVERVIEW.....	16
3.2	DATA CONVERSION PROCESS.....	16

Document Control

Version Number	Description	Release Date	Author
1.0	Initial Issue	July 19, 2002	Randy Zagorin

1 Introduction

1.1 Objective

This document describes how data will be converted from the legacy National Student Loan Data System (NSLDS) to the reengineered NSLDS II. The Data Conversion and Migration Strategy will define the NSLDS II data conversion approach, strategy, as well as the data conversion process that will be used to populate NSLDS II with data. This document will be used to communicate the data conversion approach to the impacted business owners of NSLDS II. The conversion team will investigate the conversion requirements and work with the relevant functional teams to ensure that the required data is identified and converted into NSLDS II.

1.2 Background

NSLDS is a comprehensive repository of information about Title IV Aid recipients and their loans, grants, lenders, Guaranty Agencies (GAs), servicers, and schools. NSLDS provides an integrated view of Title IV loans and grants during all stages of their life cycle from aid approval through disbursement, repayment, delinquency, and closure.

The legacy NSLDS is an IBM DB2 database stored on an IBM 9672-R85 mainframe and includes transaction-level detail on loans and default rates, with referential relationships to students, GAs, lenders and school entities. The legacy NSLDS is a classic Online Transaction Processing (OLTP) design and includes normalized data structures in the 3rd normal form.

NSLDS is comprised of multiple databases created to serve different purposes.

- Active Database – Designed to hold loan-level information for open student/borrower accounts.
- Archive Database – Designed to hold loan-level information for closed student/borrower accounts.
- Online Statistical Abstract (STAB) Database – Created in response to research teams requesting a random and statistically valid sample of NSLDS data.

While the Active Database was designed to hold open loan-level information, in reality it contains both open and closed loan-level information for student/borrower accounts. The Archive Database has never been populated with data.

1.3 Scope

The data conversion will extract, transform, and load NSLDS data from the Active Database and STAB Database into NSLDS II. The data conversion process will be composed of phases that will require a design, development, and unit testing effort. Once the individual phases are developed and unit tested, a system test of the entire conversion process will occur.

All loan-level information for open and closed student/borrower accounts will be converted from the NSLDS Active Database to NSLDS II. STAB Database data will also be converted into NSLDS II. The legacy Archive Database will not be converted, since it is not populated with data. As part of the NSLDS II data conversion, a strategy for archiving loan-level information for closed student/borrower accounts will be implemented, which will require data to be loaded into the NSLDS II Archive database during the conversion of the NSLDS Active Database.

To maintain data integrity during conversion, data validation checks will be performed at several steps in the process. At each data validation point, data will be reconciled against the data from the previous step.

The system test will involve the extraction, transformation, and load of actual data into NSLDS II. The intention of the system test is to uncover any errors in the process caused by data anomalies in the actual data and to establish time baselines for moving the data into NSLDS II. The final conversion schedule will be documented during the system test and will include the expected impact to normal FSA operations caused by the data conversion.

1.4 Conversion Architecture

An Extract, Transform, and Load (ETL) Tool will be used to manage the conversion of NSLDS data to NSLDS II. Informatica PowerCenter 5.1 will be used as the ETL tool for the NSLDS data conversion.

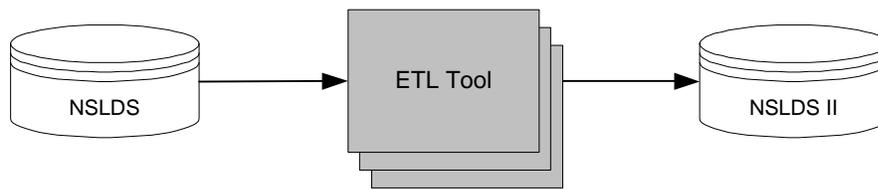


Figure 1, High Level Data Conversion Process

The ETL tool administers the selection of relevant data records for extraction as well as cleanses the extracted data for content, verifies the context, and corrects the data using business rules. The NSLDS data is loaded into NSLDS II according to the mapping established in the ETL tool.

Informatica PowerCenter 5.1 is an object-oriented, visual data transformation engine that includes the tools required to manage the detailed functions that make up the ETL process. PowerCenter 5.1 provides tools for designing, implementing, and managing the data extraction process then transforming the data and loading it into the target system (NSLDS II). The NSLDS mainframe is located at the FSA Virtual Data Center (VDC). A new instance of the NSLDS database containing the required Active database and STAB database tables to be converted will be generated prior to the conversion.

Table data from this new instance will be transferred electronically between NSLDS and the transformation server and the transformation server and NSLDS II. The bandwidth requirements for the file transfers will be identified during Detailed Design and the requisite bandwidth secured prior to conversion. Lastly, the file structures required to load the data into the transformation server and into NSLDS II will also need to be identified.

2 Data Conversion Approach

2.1 Overview

NSLDS is updated daily through both internal and external interfaces as well as through the NSLDS Financial Aid Professional Website, CICS, and by the Raytheon Quality Assurance Team. The NSLDS data conversion will be broken into two separate conversions to ensure that all updates to NSLDS are included. The initial conversion will be comprised of NSLDS data prior to a specific date, or conversion date, that has been copied as a new instance of the NSLDS database. The second conversion is actually a series of conversions that will load the interface files received post-conversion into NSLDS II. During the second conversion, information that has been updated through the NSLDS Financial Aid Professional Website, CICS, or by the Raytheon Quality Assurance team will be updated in NSLDS II as well.

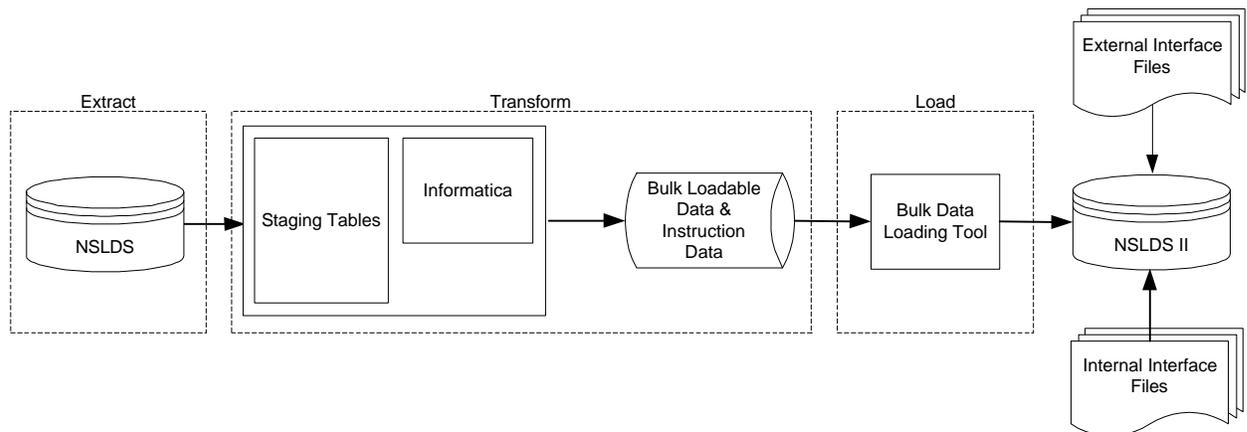


Figure 2, Conversion Process

The initial conversion process will be separated into three steps: extract, transform, and load as shown in the figure above. At several points during the ETL process the data will be verified for accuracy in NSLDS II. The secondary conversion will use the NSLDS II interface modules to load and verify the internal and external interface data along with the online updates.

2.2 Initial Data Conversion

2.2.1 Data Extraction

Data extraction is used to obtain the legacy NSLDS data from the mainframe. Specific data elements will be identified through analysis of the new data model and business requirements. The NSLDS Active database includes core tables that store the factual data as well as working tables that store operational data. The working table data is required for the general operation of the legacy NSLDS database. The core tables will be converted to NSLDS II, while working tables will not. The extraction will be performed on an instance of the NSLDS database generated specifically for the data conversion. The conversion instance of NSLDS will be

compromised of the following tables from the Active and STAB databases. This list will be finalized, as the NSLDS II Logical Data Model is refined during Detailed Design.

Table Description	Table Name	Database	
Aid Overpayment	AID_OVRPMT	Active	
Appeal Request	APPEAL_REQ	Active	
Anticipated Completion	ANTIC_COMPL	Active	
Certification	CERT	Active	
Collection	COLL	Active	STAB
Deferment Rate History	DEF_RT_HIS	Active	
Deferment Type	DFR_TYPE	Active	STAB
Enrollment	ENRL	Active	
Expected Student Enrollment	EXP_STU_ENRL	Active	STAB
FDLP Servicer	FDSL_P_SVR	Active	STAB
FDLP Servicer Branch Holder	FDSL_P_SVR_BR_HOL	Active	STAB
FDLP Servicer Submittal History	FS_SBMTL_HIS	Active	
FDLP Servicer Submittal Run Error	FS_SBMTL_RUN_ERROR	Active	
Financial Profile	FINANC_PROF	Active	STAB
Guaranty Agency	GA	Active	STAB
GA Aggregate	GA_AGG	Active	
GA Summary	GA_SUM	Active	
GA Submittal History	GA_SBMTL_HIS	Active	
GA Submittal Run Error	GA_SBMTL_RUN_ERR	Active	
Insurance Claim Payment	INSUR_CL_PMT	Active	STAB
Insurance Claim Refund	INSUR_CL_RFD	Active	STAB
IRS Offset	IRS_OFF	Active	STAB
Lender	LEN	Active	STAB
Lender Branch	LEN_BR	Active	STAB
Lender Branch Aggregate	LEN_BR_AGG	Active	
Lender Branch Holder	LEN_BR_HOL	Active	STAB
Lender Branch Holder Servicer	LEN_BR_HOL_SVR	Active	STAB
Lender Branch Servicer	LEN_BR_SVR	Active	STAB
Lender Branch Servicer Aggregate	LEN_BR_SVR_AGG	Active	
Loan	LOAN	Active	STAB
Loan Cancellation	LOAN_CAN	Active	STAB
Loan Default Rate Default	LOAN_DEF_RT_DEF	Active	
Loan Deferment	LOAN_DFR	Active	STAB
Loan Disbursement	LOAN_DIS	Active	STAB
Loan External Identification	LOAN_EXTL_ID	Active	
Loan Forgiveness	LOAN_FRGV	Active	

Table Description	Table Name	Database	
Loan Guarantor	LOAN_GUA	Active	STAB
Loan Origination Support	LOAN_ORIGN_SUPP	Active	
Loan Refund	LOAN_RFD	Active	STAB
Loan Repayment Plan	LOAN_RPMT_PLAN	Active	STAB
Loan Status	LOAN_STAT	Active	STAB
Loan Type	LOAN_TYPE	Active	STAB
Loan Type Group	LOAN_TYPE_GP	Active	
Loan Unpaid Refund Discharged	LOAN_UNPD_RFD_DCHG	Active	
National Aggregate	NATL_AGG	Active	
Organization	ORG	Active	
NSLDS User	NSLDS_USER	Active	
PCA SPA Notification	PCA_SPA_NOTIF	Active	STAB
Pell Grant	PELL_GRT	Active	STAB
PLUS Borrower	PLUS_BOR	Active	STAB
Portfolio Status	PORT_STAT	Active	
Prescreening Result	PRSCRN_RSLT	Active	
Prescreening Result Loan	PRSCRN_RSLT_LOAN	Active	
Region Aggregate	REG_AGG	Active	
Reinsurance Claim Payment	REINSUR_CL_PMT	Active	STAB
Reinsurance Claim Refund	REINSUR_CL_RFD	Active	STAB
Repurchased Loan	RPCH_LOAN	Active	STAB
School	SCH	Active	STAB
School Branch	SCH_BR	Active	STAB
School Branch Aggregate	SCH_BR_AGG	Active	
School Branch Holder Servicer	SCH_BR_HOL_SVR	Active	STAB
School Branch Status History	SCH_BR_STAT_HIS	Active	
School Branch Servicer Aggregate	SCH_BR_SVR_AGG	Active	
School Branch Transmittal History	SCH_BR_TRAN_HIS	Active	STAB
School Origination History	SCH_ORIGN_HIS	Active	STAB
School Submittal History	SCH_SBMTL_HIS	Active	
School Submittal Run Errors	SCH_SBMTL_RUN_ERR	Active	
SPA Payment	SPA_PMT	Active	STAB
SSCR Performance	SSCR_PERF	Active	
SSCR Regeneration	SSCR_REGEN	Active	
State Aggregate	ST_AGG	Active	
Student	STU	Active	STAB
Student Address	STU_ADD	Active	
Student Branch Identifier	STU_BR_ID	Active	
Student Name	STU_NM	Active	STAB

Table Description	Table Name	Database	
Submittal Error Description	SBMTL_ERR_DESCN	Active	
Supplemental Reinsurance Payment	SUPP_REINSUR_PMT	Active	STAB
Transfer Monitoring Student	TM_STU	Active	
Transfer Student Alert	TM_STU_ALERT	Active	
Transfer Monitoring Student Alert Loan	TM_STU_ALERT_LOAN	Active	
Transfer Monitoring Student Alert Name	TM_STU_ALERT_NM	Active	
Transfer Monitoring Student Alert Overpayment	TM_STU_ALERT_OVRPMT	Active	
Transfer Monitoring Student Alert Pell	TM_STU_ALERT_PELL	Active	

Table 1, NSLDS Active and STAB database tables to be converted to NSLDS II

There are two methods available for extracting NSLDS data:

- Informatica PowerConnect - Informatica PowerCenter 5.1 would be integrated with NSLDS using Informatica PowerConnect for Mainframe. PowerConnect would allow for the extraction of data directly from NSLDS.
- Flat files – NSLDS data would be extracted to flat files.

During detailed design, the method for extracting NSLDS data and the specific data entities along with attributes in NSLDS that will be used to populate NSLDS II will be finalized.

Prior to beginning the data conversion, the volume of each table and complexity of the data structures will be documented. Using this information, a detailed mapping table will be developed for the NSLDS data elements, which will specify how the data will be moved from NSLDS to NSLDS II. The functionality of the data, data type, length, constraints, and relationships with other tables will also be taken into account during the mapping process. Any inconsistencies between the NSLDS and NSLDS II data models will be noted and processing rules, translation rules, and default values will be determined to resolve data mapping gaps. During the development of the detailed mapping table, the NSLDS II Reengineering team will take into consideration the Single Sign On, Common School ID, and Common Student ID efforts underway at Modernization Partner and the effect that these initiatives may have on NSLDS II. The finished mapping table will be used to determine the data elements that will be extracted.

The extraction module will be built/configured and unit tested. The unit testing will be comprised of a combination of data selects, file formatting, and file transfers that will be used to verify the data being extracted is accurate and in the correct format.

2.2.2 Data Transformation

Transformation is used to cleanse and integrate the extracted data. Data will be received from the extract step and enhanced prior to being loaded into NSLDS II based upon the requirements outlined in the mapping tables. The ETL tool will be used to manage this transformation process. Transformation will be performed on a dedicated transformation server.

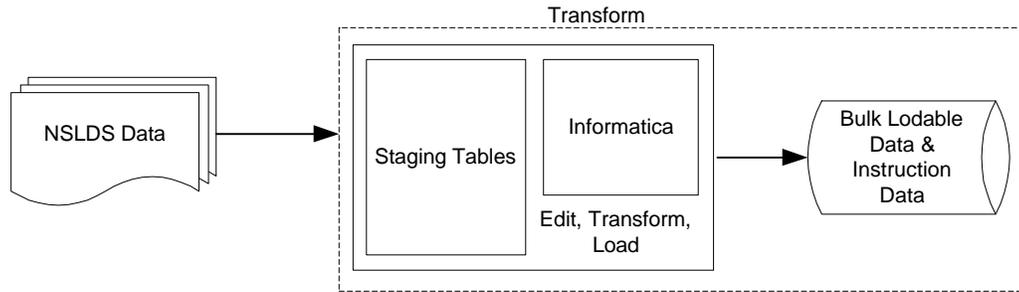


Figure 3, Transformation Process

The ETL Tool will be used to load the data into staging tables on the transformation server as shown in the figure above. Staging refers to a temporary holding area for the data prior to undergoing transformation. The staging table layouts will be developed and documented during detailed design. The mapping tables will form the basis of the transformation design and the transformation logic will be developed using the ETL Tool. During the transformation, data will be cleansed, mapped, and aggregated.

The data will then be examined to determine which elements should be cleansed. Cleansing includes editing the data for content, checking the context of the data, and correcting data according to business rules. NSLDS data will only be cleansed if the data source can be updated so subsequent updates do not inaccurately overwrite the cleansed data with incorrect information. Cleansing may also be limited to a specific period of time depending on business rules and the data used.

The mapping of the data is the process of sending NSLDS data to the proper location in NSLDS II according to a one-to-one relationship between the source and target database. Each piece of information brought into NSLDS II must be sent to a specific location. Mapping is the process of defining where the data should go and includes matching, elementization, and standardization.

- Matching is the process of checking for similar data elements and prevents redundant storage, indexing, and searching.
- Elementization is parsing or separating data into a list of standard elements.
- Standardization is arranging data in a single, consistent format.

The transformation process will also include the aggregation of data, which is the process of creating higher-level roll-up records from the detailed source records. Finally, the data will be formatted for bulk loading into NSLDS II. The transformation logic, data cleansing, and data aggregation will be documented during detail design.

Data that fails to load into NSLDS II will be noted in an exception report. These data elements will be inspected to determine the reason that they were not processed and then repaired. The repaired data will be sent back through the transformation process. If the repairs to the data were effective, the data will go through cleanly. Otherwise, the data will be written back to the exception file, reviewed, and repaired again.

The build and test phases of the data transformation process include building the staging tables, loading the staging tables, executing the conversion procedures, and generating the validation reports. The staging tables will be built and maintained on the transformation server based upon the staging table layouts. Unit testing will verify that the extracted data can be populated in the staging tables.

The conversion procedures will be used to establish the transformation logic. The procedures will be built to select, clean, and convert the NSLDS data. Unit testing will compare the data before it has been transformed to the data after the transformation process has been completed.

2.2.3 Data Load

During the data load, data will be received from the transform step and loaded into NSLDS II. The data load will be accomplished using a specialized load utility that directly loads the data from the source files produced by the ETL Tool during transformation.

The conversion procedures, mapping tables, and new data model will form the basis of the load design. The NSLDS II tables will be populated based upon the requirements outlined in the conversion procedures and mapping tables. The logic required for the load tool to insert the data will be developed as well.

The build and test phase of the data load will include the loading of NSLDS II tables as well as the development of validation reports. The ETL tool will generate files that are formatted for the load utility. The load utility will move the data into NSLDS II. Unit testing will compare the data in NSLDS II to the data on the transformation server.

2.3 Secondary Data Conversion

The data received by NSLDS after the conversion date will be loaded into NSLDS II following the initial conversion. The standard NSLDS II interface load modules will be used to load this data. Until all NSLDS data is converted during the initial conversion, incoming interface files will be held in a queue. Once the initial conversion is complete, the interface files will be loaded into NSLDS II. When possible the files in the queue will be loaded in parallel instead of sequentially to reduce the processing time.

External interface files that are sent electronically to NSLDS are received in a mailbox that is accessible through the Student Aid Internet Gateway (SAIG) using the EDconnect software and/or Online Secure Manager. Online Secure Manager allows users to access an Archive of all files sent to the mailbox within the last 180 days. This list will be used to verify that all external interface files that are sent electronically have been received and to retrieve any missing files. External interface files loaded into NSLDS by tape must be identified prior to the conversion. Finally, the process for collecting and identifying interface files from internal systems that will send updates to NSLDS II through the EAI Bus will also need to be identified.

The performance implications of using the interface modules to load interface files for conversion will have to be investigated and tested. Where performance issues are identified, modified versions of the interfaces may need to be developed. After the interface files are loaded in NSLDS II the data will need to be validated.

The secondary data conversion will also include the manual updates made to NSLDS data from the following sources:

- Financial Aid Professional web site
- CICS
- Raytheon Quality Assurance Team

Manual updates made to NSLDS after the conversion begins will be tracked and logged so that the updates may be made in NSLDS II as well. Log files that include a date and time stamp as well as the updated data will be used to capture the updates made through the Financial Aid Professional web site and CICS.

The log files for the Financial Aid Professional web site will be generated using the middleware application that interconnects the web server with the NSLDS mainframe. The log files for updates made through CICS will be gathered from the mainframe. Additionally, the list of updates made to NSLDS by the Raytheon Quality Assurance Team will be generated from their SIR log and/or change log.

2.4 Aggregate Tables

NSLDS currently stores aggregate data on schools, lenders, GAs, and servicers as well as each ED region, state, and nationally in the Active Database. Aggregate data is calculated internally and represents a summarization of data from NSLDS. This data is used in NSLDS to support users by reducing processing times and improving performance.

NSLDS has eight tables where aggregate data is stored, which are produced monthly, quarterly, or annually. Monthly aggregates are run 15 days after the end of the month while quarterly and yearly aggregates are run 45 days after the end of the quarter/year. The current data model provides the flexibility to create aggregates for prior time periods if an aggregate needs to be recreated. Aggregate data is retained online for 18 months.

The aggregate tables in NSLDS are as follows:

- SCHOOL_BRANCH_AGGREGATE—Populated monthly
- SCHOOL_BRANCH_SERVICER_AGGREGATE—Populated monthly
- GA_AGGREGATE—Populated monthly and quarterly
- LENDER_BRANCH_AGGREGATE—Populated quarterly
- LENDER_BRANCH_SERVICER_AGGREGATE—Populated quarterly
- STATE_AGGREGATE—Populated monthly and yearly
- REGION_AGGREGATE—Populated monthly
- NATIONAL_AGGREGATE—Populated monthly and yearly

Aggregate data will be extracted from NSLDS during the conversion of the Active Database to NSLDS II and will be maintained online in NSLDS II for 18 months. This data will be converted to ensure that users are able to access aggregate data from time periods prior to the conversion in NSLDS II. To ensure that the data conversion has been successfully completed, aggregate data will be generated in NSLDS II and validated against the NSLDS aggregate data.

2.5 Data Validation

Data validation will occur at numerous milestones within the conversion process. Various data validation techniques will be built into the data conversion steps and standalone queries and reports will be used to validate the converted data. Validation techniques will also include the following:

- Verifying the total record counts of the target tables and the source tables (e.g. the transformation server will be the target and the NSLDS tables will be the source during the transformation step).
- Sum totals for fields will be calculated in the target tables and compared to the totals in the source tables.
- Reports will be executed on the target system and the source system and reviewed for discrepancies.

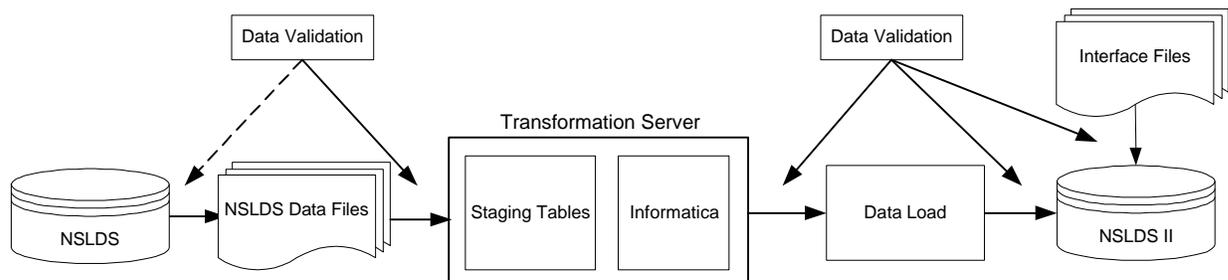


Figure 4, Data Validation Process

As illustrated above, during the initial conversion data will be validated when:

- NSLDS data is extracted into the NSLDS Data Files (This step is only required if NSLDS data is extracted to flat files.)
- NSLDS data files are loaded into the staging tables within the Transformation Server
- The data in the staging tables has been transformed and prepared for data load
- Data is loaded into NSLDS II

After the initial data conversion is complete, the data in NSLDS II will be validated against data in NSLDS. During the secondary conversion data will be validated when the interface files are loaded into NSLDS II. After all interface files have been loaded into NSLDS II, the data will once again be validated against data in NSLDS. At this point the legacy NSLDS database and the reengineered NSLDS II database will be synchronized.

2.6 Archive Strategy/Conversion

As part of the original design of NSLDS an Archive Database was created. This database was designed to hold loan-level information for closed loans student/borrower accounts. The Archive Database was never populated with data and is currently empty; therefore, a data conversion is not necessary.

Although the NSLDS Archive Database was never populated with data, an Archive Database will be established and populated in NSLDS II. During the NSLDS II data conversion process, loan-level information for certain closed student/borrower accounts will be archived.

2.7 Online Statistical Abstract Conversion

NSLDS contains an Online Statistical Abstract (STAB) database that is created or refreshed quarterly. The database is used for statistical and research purposes and contains a sanitized subset of NSLDS data comprised of a random sampling of 1.5 million borrowers from the Active Database. The STAB database tables will be converted to NSLDS II during the initial data conversion.

The STAB database in NSLDS II will be validated against the STAB databases in NSLDS to ensure the data conversion was successful by:

- Verifying table row counts
- Comparing the sum total calculated fields
- Verifying aggregate data
- Verifying a random sampling of the STAB data

2.8 System Test

System testing allows for high-level validation of the ETL process. The ETL process can be verified by comparing the number of rows transferred, the overall time required, and outputs generated from the target database. The primary goal is to discover and fix any errors in the data load. A secondary goal is to minimize the overall process time.

Once the initial and secondary data conversion processes have been built and unit tested, several conversion test runs will be performed as part of the system testing effort. During each of the test runs, the size of the data set will be increased until the entire NSLDS database has been converted during a single test run. After each run is completed the converted data in NSLDS II will be verified with the data in NSLDS. The system test will also be used to develop a baseline of the time required to complete the entire data conversion

The initial system test run will be performed on a partial extract of the data to ensure that the process is working correctly. A cross section of data will be examined in each system to ensure that the transformation and load processes were successful. Additionally, row counts will be validated to ensure that all of the transformed data was loaded into NSLDS II. Next the volume

of the data in the sample will be increase incrementally in the subsequent test runs as the results from the previous runs are verified. Following the completion of a successful system test on a partial extract of data, the NSLDS II test database will be created.

At this time, the preliminary test runs will not include the loading of the interface files that are part of the secondary conversion. The last step in the overall conversion system test will be to load an interface file from each of the internal and external sources. This test will be used to verify the time required to load each of the files so it can be tracked and baselined. Additionally, the data content will be verified.

The system test of the data conversion should begin a few months before the actual NSLDS data conversion is scheduled. The time required to complete the conversion and the number of records converted will be tracked to validate the estimated conversion schedule. The final conversion schedule will be developed and will document the expected impact to normal FSA operations caused by the data conversion.

2.9 Assumptions

- Key resources from across the functional areas and from Raytheon will be available during the design and testing processes.
- Informatica PowerCenter 5.1 will be selected as the ETL Tool.
- The Informatica PowerCenter 5.1 software will be available prior to the data conversion build.
- Informatica PowerConnect will be used to access legacy NSLDS data.
- Raytheon will create an instance of the NSLDS Active and STAB database that contains the NSLDS tables to be converted prior to the conversion.
- The interface files loaded into NSLDS after the new instance has been generated will be available and stored for loading into NSLDS II.
- A method for accessing archived data from NSLDS II will be available and NSLDS data for closed loan-level student/borrower information will be archived in NSLDS II.
- The interface modules will be used to load data into NSLDS II from external and internal interfaces during the data conversion.
- The Modernization Partner NSLDS II team will work with the legacy NSLDS team and the FSA hosting site to track and record the updates made to the legacy NSLDS through the NSLDS Financial Aid Professional web site, CICS, and by the Raytheon Quality Assurance Team during the initial conversion cycle.

2.10 Risks and Contingencies

Risk	Contingency
Key resources from across the functional areas and from Raytheon will be available during the design and testing processes.	The data migration testing and conversion process will require key resources from across the functional areas and from Raytheon. These resources will be identified and secured during detailed design.
Informatica PowerConnect will be used to access legacy NSLDS data.	Legacy NSLDS data will be extracted to flat files that will be loaded into Informatica.
A data archiving strategy for NSLDS data needs to be developed and approved.	Key resources from across the functional areas will need to approve the archiving strategy during detailed design.
An instance of the NSLDS Active and STAB database that contains the NSLDS tables to be converted needs to be created by Raytheon prior to the conversion.	The conversion team will begin addressing this risk during detailed design to ensure that the process for creating the instance has been established and is agreed upon by Raytheon.
Updates made to the legacy through the NSLDS Financial Aid Professional web site, CICS, and by the Raytheon Quality Assurance Team during the initial conversion cycle NSLDS need to be tracked and recorded.	The Modernization Partner NSLDS II team will work with the legacy NSLDS team and the FSA hosting site to track and record the updates made to the legacy NSLDS through the NSLDS Financial Aid Professional web site, CICS, and by the Raytheon Quality Assurance Team during the initial conversion cycle.

Table 2, Risks and Contingencies

3 Data Conversion Plan

3.1 Overview

The NSLDS conversion process involves converting data in NSLDS into the NSLDS II data model. The conversion will include NSLDS data stored in the Active Database as well as the STAB database. Both data conversions will follow the same process, except that a secondary conversion will not be required for the STAB database. Closed loan-level information stored in the NSLDS Active database for loan-level student/borrower accounts will be archived to the NSLDS II Archive Database during the conversion.

The conversion process will be broken down into the following steps:

- Step 1 – Create A New Instance of the NSLDS Database
- Step 2 – Extract Data From NSLDS
- Step 3 – Reconcile Extract Files Against NSLDS Data
- Step 4 – Transmit Flat Files To The Transformation Server
- Step 5 – Load Data Into The Staging Tables
- Step 6 – Reconcile Loaded Record Count Against Extracted Record Count
- Step 7 – Transform The Data
- Step 8 – Verify All Data Was Transformed
- Step 9 – Load The Data Into NSLDS II
- Step 10 – Reconcile The Data Loaded Into NSLDS II
- Step 11 – Secondary Conversion

The steps are described in detail below.

3.2 Data Conversion Process

Step 1 – Create A New Instance of the NSLDS Database

Raytheon will create a new instance of the legacy NSLDS database containing the core tables and other tables that will be converted from the Active database and STAB database. The new instance will be stored on the mainframe. The extraction will be performed on this instance of the NSLDS database. During the creation of the new instance for conversion, no updates will be made to the legacy NSLDS.

Step 2 – Extract Data From NSLDS

Data will be extracted from the instance of the NSLDS Active Database, which resides on the mainframe at the VDC, in a format that is compatible for loading into the staging tables. NSLDS data will first be extracted from the Active Database, which includes the NSLDS core tables and the aggregate tables. Once the Active Database extract is complete, the STAB Database tables will be extracted. A report on the number of records in each table will be produced allowing the extract to be reconciled against the data in NSLDS.

Step 3 – Reconcile Extract Files Against NSLDS Data

NSLDS data extracted to flat files, must be verified to ensure that is consistent with the data stored in NSLDS before it can be loading into the staging tables on the transformation server. The actual row counts from NSLDS tables will be compared against the row counts in the flat files to ensure that the NSLDS data had been properly extracted.

If Informatica PowerConnect for Mainframes is used to extract the data from NSLDS, data will be loaded directly into the staging tables on the transformation server and will be reconciled after the staging tables are loaded.

Step 4 – Transmit Flat Files To The Transformation Server

Once the NSLDS data is extracted to flat files, the data will be transmitted electronically via ftp or through the EAI Bus to the transformation server.

If Informatica PowerCenter and PowerConnect for Mainframe are used to extract NSLDS data, this step would not be applicable since PowerConnect establishes a direct link between NSLDS and the transformation server.

Step 5 – Load NSLDS Data Into The Staging Tables

NSLDS data will be loaded into the staging tables on the transformation server. Data will be loaded using parallel processing, which allows large data to be split into smaller groups and loaded concurrently. Checkpoint restarting will be used to designate points during the load were the process can be restarted if the data load fails, to mitigate the risk of needing to restart the load process from the beginning should a problem arise.

Statistics about the load session, such as the number of records read and written, will be tracked. This information will be stored in log files and will be used to verify that all the records were processed properly. Additionally, the number of rows accepted and rejected during the load process will be recorded. All rejected records will be reviewed and any issues will be resolved before the next process step occurs.

Step 6 – Reconcile Loaded Record Count Against Extracted Record Count

The number of records successfully loaded into the staging tables will be compared against the number of records contained in the NSLDS data extract to ensure that all of the data has been loaded.

Step 7 – Transform the Data

Once the data has been loaded into the staging tables on the transformation server, data will be converted from the NSLDS data model into the NSLDS II data model. The transformation will include identifying redundant data, cleansing data, aggregating data, and mapping the data to

the appropriate location in NSLDS II. Transformed records will be recorded in a transformed records file. Any record that cannot be transformed will be written to a rejected records file. The rejected records file will be used to determine the reason the data failed to be transformed. The rejected records will be manually corrected and the data will be validated against legacy NSLDS data. The rejected records will be loaded back into the target database after correcting the error. Additionally, loan-level information for closed student/borrower accounts will be identified and mapped to fields in the NSLDS II Archive Database.

Once the transformation has been completed, data will be prepared for bulk loading into NSLDS II.

Step 8 – Verify All Data Was Transformed

The rejected records file will be reviewed to ensure that all records were transformed and prepared for loading. Additionally, the total number of records transformed will be verified against the total number of records prepared for bulk loading.

Step 9 – Load the Data Into NSLDS II

Data will be sent from the transformation server to NSLDS II. The data load will be accomplished using a load utility that loads the transformed data directly into NSLDS II. The target fields and tables will be defined based upon the mapping tables. Data that fails to be loaded into NSLDS II will be recorded, corrected, and reloaded.

Step 10 – Reconcile the Data Loaded Into NSLDS II

This is the final reconciliation in the conversion process, which is intended to ensure that all the data that was transformed is loaded. Row counts from NSLDS II will be reconciled against the transformed data to ensure all the required data was loaded into NSLDS II.

Step 11 – Secondary Conversion

Interface files received by NSLDS after the conversion date will be loaded into NSLDS II using modules developed by the interface team. The list of interface files will be verified against the list of files received in the NSLDS SAIG mailbox and list of files received via tape. Additionally, interface files from internal systems that were sent to NSLDS II through the EAI Bus will also be loaded. Interface files will be loaded in chronological order. When possible, interface files will be loaded in parallel to reduce the conversion time.

After the internal and external interfaces files are loaded for a specific day, the data manually updated through the NSLDS Financial Aid Professional web site, CICS, or by the Raytheon Quality Assurance team, will be loaded into NSLDS II. Once this information has been loaded, the interface load for the next day will begin. This process will be repeated until all of the interface files and all of the manual updates have been entered into NSLDS II.