

**DEPARTMENT OF EDUCATION-
FEDERAL STUDENT AID
FINANCIAL PARTNERS**



**INTERFACE DESIGN
FUNCTIONAL DESIGN DOCUMENT
FINANCIAL PARTNERS DATA MART RELEASE III**

July 18, 2003
Version 1.0

TABLE OF CONTENTS

<i>1 DATA ACQUISITION AND DATA MAPPING</i>	<i>1</i>
1.1 INTERFACE DESIGN	1
1.2 INTERFACE PROCESS	1
<i>1.2.1 FMS</i>	2
<i>1.2.2 NSLDS</i>	2
<i>1.2.3 PEPS</i>	2
1.3 ERROR HANDLING	3
1.4 DATA QUALITY ASSURANCE	3
<i>2 DATABASE DESIGN</i>	<i>5</i>
2.1 LOGICAL DATA MODEL	5
2.2 PHYSICAL DATABASE	6
<i>3 TECHNICAL ARCHITECTURE</i>	<i>8</i>
3.1 OVERVIEW OF TECHNICAL ARCHITECTURE	8
3.2 SOFTWARE	10
3.3 HARDWARE	10
3.4 MIDDLEWARE	11

1 Data Acquisition and Data Mapping

This document describes how the data will be received, staged (into a staging area), transformed and loaded into the FP Data Mart. It also describes the data quality assurance and error handling processes. Data acquisition typically uses COTS tools, which specialize in this function. This document will also provide as an appendix the data mapping from the NSLDS, FMS, and PEPS systems to the FP Data Mart.

1.1 Interface Design

Data acquisition, which is also known as extract-transform-load (ETL) consists of three steps:

1. **Extraction** - is the process of acquiring data from one or more systems for the purpose of loading a Data Mart. In the FP Data Mart architecture, NSLDS, FMS, and PEPS will send one or more files on a monthly basis to the Data Mart's staging area. Informatica will then work against this data to validate and cleanse it before loading into the FP Data Mart.
2. **Transformation** - is a general term for cleansing and validating incoming data, which includes: handling missing elements, looking up tables, aggregating rows and standardizing field formats. In the FP Data Mart architecture, Informatica will handle these types of transformations and field-to-field mappings.
3. **Loading** - there are two ways to load data: record-by-record through the Informatica Server Manager or through the Oracle Bulk Loader utility. In the FP Data Mart architecture, we will only be adding data, never updating or deleting specific records. Therefore, we will use the Oracle Bulk Loader utility, as necessary, to load the fact tables.

1.2 Interface Process

The FMS, NSLDS, and PEPS systems were implemented as source systems for the second release of the FP Data Mart and will continue to provide the existing data feeds, along with new data that was identified for the third release. Based upon the end user requirements, queries/extract programs have been designed and the corresponding data will be extracted from the source systems. These programs will produce sequential files that will be made available to the EAI architecture. The files will then be transmitted to the Informatica server and will be used by the Informatica ETL process.

Some of the data to support new 'fact' or 'dimension' tables will be extracted from the source systems or created using a spreadsheet. This data will then be loaded directly into the Data Mart tables and continue to be updated on a monthly basis if necessary.

Some modifications will be made to Informatica Mappings to incorporate any changes in data feed structure or data feeds that have gone away since the first release of FP Data Mart. It has been concluded that reports using this data will have little or no affect due to the removal of these data feeds, however modifications will be made as necessary if needed.

1.2.1 FMS

Data that has been added or changed since the last execution of the extract programs/queries will be extracted and transmitted using the EAI architecture to the Data Mart's Informatica server. The extract is expected to occur on a monthly basis. The existing extract will continue to obtain data related to Guaranty Agencies (Form 2000, VFA) and State Agencies (LEAP/SLEAP) from FMS. The new extract identified in this release will provide the LaRS data feed to replace the LENDER feed for Form 799 and shall include all historical data since the Raytheon feed was last received in September of 2002. Data from the Mellon Interface will be extracted to provide Lender Fee Payment information and Lender address information will be obtained from FMS as well. Only the most recent and up to date information that is available in FMS will be extracted every month. The data will continue to be delivered via the EAI bus on a monthly basis, based on the last Sunday of the month.

For a list of new tables and columns that will be extracted for this release, reference Appendix A: Source to Target Mapping.

1.2.2 NSLDS

In the second release, FP Data Mart identified the data related to student loans associated with Lenders, Guaranty Agencies, Servicers, and Schools from NSLDS would be extracted. The existing data extracted from NSLDS is a 'snapshot' of the data instead of extracting only the 'changed' data. The 'snapshot' will continue to be extracted monthly and transmitted using the EAI architecture to the Data Mart's Informatica server. In the third release, it was determined that three additional data extractions would be needed. This included year-end summary information for Cohort Default Rate values, Total Dollars in Subrogated Loans, and AMF (Account Maintenance Fee) information for GAs. The first data load will include historical data that has been backed up to tape. These extractions will then continue to be made monthly or at fiscal year end, and will be transmitted along with current data files via the EAI Bus.

For a list of new tables and columns that will be extracted for this release, reference Appendix A: Source to Target Mapping.

1.2.3 PEPS

Data that has been added or changed since the last execution of the extract programs/queries will be extracted and transmitted using the EAI architecture to the Data Mart's Informatica server. The extract will continue to occur on a monthly basis. The existing extraction from PEPS included data related to Lender audits, Lender program reviews, and School closings from PEPS. Going forward the PEPS system will also provide Lender Association information that was previously received from Financial Partners for the first release of the Data Mart. This will ensure that data is being obtained from the source system to provide quality assurance of data. This data extract will be used in conjunction with the demographic data that will be obtained from LaRS to replace the ALLLIDS.txt file. This data will be extracted on the last Sunday of the month.

For a list of tables and columns that will be extracted, reference Appendix A: Source to Target Mapping.

1.3 Error Handling

The error handling and validation processes have been designed to provide the users with as much source system data as possible, that is, to load the target tables with the same data as the source systems. The challenge here is that the source systems have different data models from the Data Mart model. The error handling is designed to provide load level validation, that is, those rows which do not conform to the rules of *referential integrity** will be rejected (e.g., duplicative key fields, etc.). These rows will be written to an error file on the Informatica server. These error files can be viewed by subject matter experts to reconcile the error records.

**Referential Integrity - A feature is provided by relational database management systems (RDBMS) that prevents users or applications from entering inconsistent data. Most RDBMS have various referential integrity rules that can be applied when creating a relationship between two tables. For example, referential integrity would prevent adding a record to Table B that cannot be linked to Table A.*

1.4 Data Quality Assurance

The quality assurance steps listed below will help to ensure that complete and correct data is loaded to the FP Data Mart.

- ◆ Use the EAI architecture to transfer files from the source systems to the Informatica server. The EAI architecture will guarantee successful delivery of the file from the source system to the Informatica server thus reducing the requirement for manual verification of the number of records on the file in the source system to that on the Informatica server.
- ◆ Verify whether the incoming data has correct values: The software will check for spaces or nulls in the input columns. If there are spaces or nulls, the value 'unknown' or nulls will be assigned to the affected columns, so that the user will obtain correct and meaningful information from the FP Data Mart.
- ◆ Verify using random sampling and checking screens/queries on the source systems with corresponding reports in the Data Mart that the values are as expected.
- ◆ Verify the Informatica input file with the values in the FP Data Mart. The input file is loaded directly into a staging table in Informatica. SQL is used to sum columns in the staging tables with columns in the FP Data Mart tables.
- ◆ Verify valid keys for incoming data: Keys will be generated for those records that have invalid keys or no key values, using Informatica's Sequence Generator.

When loading the Data Mart, it is preferable when all of the identified source records are loaded into the Data Mart. If is possible however, that some types of records may enter the source system and be transferred to the Data Mart, and yet the Data Mart is not expecting that type of record. These records may enter into the source system as a result of an edit change or an enhancement that was applied to the source system. This change may not have been communicated to the personnel maintaining the Data Mart system.

When a record from the source system contains information that the Data Mart does not expect, the following will occur:

1. For fields identified as 'code/decode' fields, the Informatica process will use the value in the field from the source record and do a 'lookup' on one of the 'lookup tables'.
2. If the Informatica process finds a match, the record continues to be processed.
3. If the Informatica process does not find a match, it will add an entry to the 'lookup table'. The 'code' will be the value in the source record field. The 'decode' or description for the code will be set to 'Unknown'. The Informatica process will then continue processing the record.
5. Routinely after all monthly processes have completed on the Data Mart, a query will be executed against the 'lookup tables' to find all entries with 'decode'/description values of 'Unknown'. These situations will be further researched to determine what value should be used for the 'decode'/description and the description updated as appropriate. If it is determined that the 'code' should not have existed in the source record, then the Data Mart team will identify this situation to the owners of the source system so that they owners may appropriately address the issue.

If there are situations similar to the above where it is determined NOT appropriate to just 'add a code' to the lookup table, then this source record will be rejected and not loaded into the Data Mart. These records will be written to an error file or identified in some other way such that they can be further researched to identify the problem and the source of the problem.

2 Database Design

2.1 Logical Data Model

The logical model shown below was designed based on information gathered from the subject matter experts (SMEs) in the FP Channel and other areas within FSA.

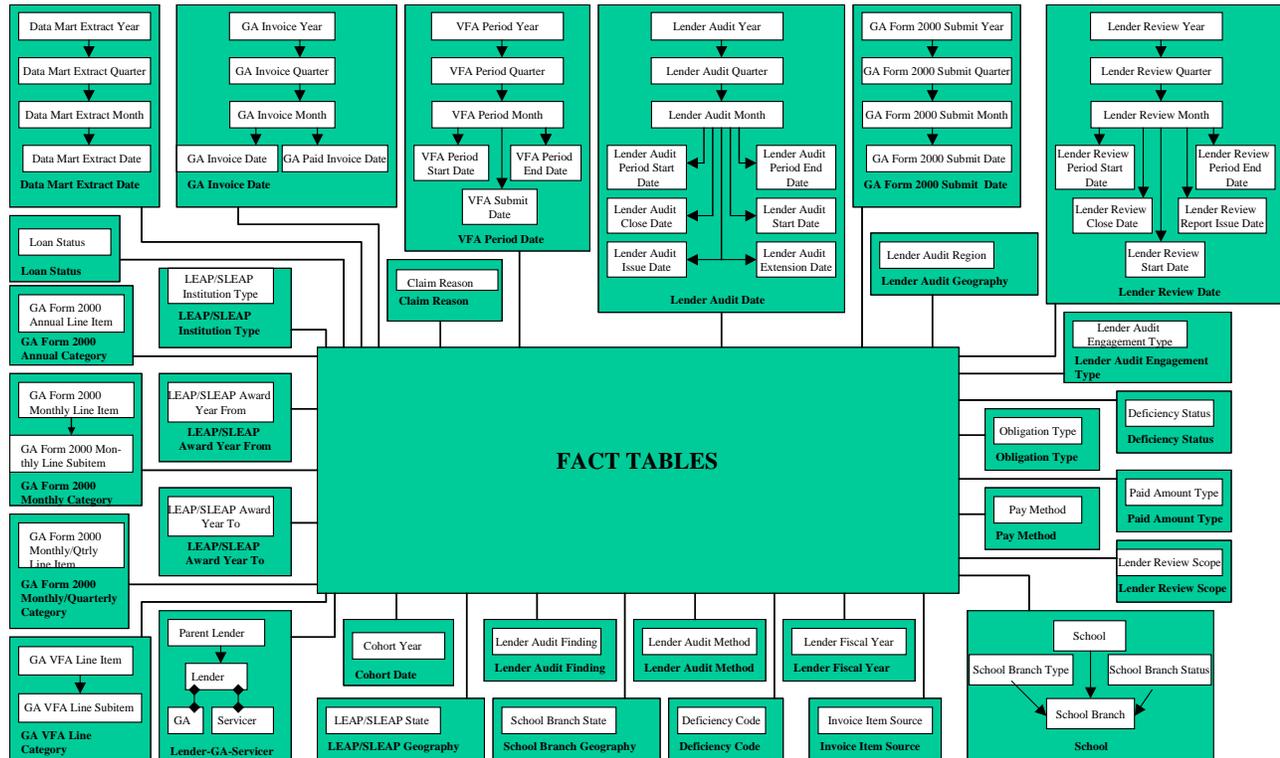


Figure 1: Release 3 Logical Data Model

The logical and physical data models for the new fact and dimension tables are available in Appendix B: Logical and Physical Data Models.

2.2 Physical Database

The design of the physical database includes the actual table name, column name, and key identification. The tables will be developed to support the identified required reports and to allow for flexibility.

The following table identifies the new Fact and Lookup tables that will be created as part of the third release for FP Data Mart.

Table Names
F_GA_AMF
F_GA_DFLT_RATES
F_GA_SCORECARD
F_GA_SCORECARD_DLT_RESULTS
F_GA_SUB_LOAN
LKP_ALLSVCRS
LKP_GA_DESIGNATION
LKP_GA_PRTFL_LVL
LKP_GA_SCORECARD_DTL
LKP_GA_SCORECARD_ITEM

Table 1: New Fact and Lookup Tables for Release 3

The logical and physical data models for the new fact and dimension tables are available in Appendix B: Logical and Physical Data Models.

For table specifications that will be affected for Release III, see the Interface Control Documents included in Appendix C.

The following table identifies the required reports and which database Fact table will be used.

Report	Fact Table
Account Maintenance Comparison	F_FM2000_ANNL F_GA_AMF
Cash Liquidity	F_FM2000_ANNL
Change in Error Rates	F_VFA_PERF_RPT
Change in Federal Funds	F_FM2000_ANNL
Change in Loan Status	F_FFEL_LOAN
Change in Operating Funds	F_FM2000_ANNL
Cohort Default Rate	F_GA_DFLT_RATES
Direct Collections Activity	F_FM2000_MNTH F_FM2000_MNTH_QTR
FFEL Collections	F_FM2000_MNTH
GA Global Summary	F_GA_SCORECARD
GA Scorecard	F_GA_SCORECARD
GA Scorecard Analysis Report	F_GA_SCORECARD
GA Subrogated Loan Candidate	F_FFEL_LOAN

Report	Fact Table
	F_GA_SUB_LOAN
History of Total Collections on Defaulted Loans	F_FM2000_MNTH F_FM2000_MNTH_QTR
Market Share Analysis	F_FFEL_LOAN
Minimum Reserve Level	F_FM2000_ANNL
Percent of Change in Default Portfolio	F_FM2000_ANNL
Portfolio Characteristics	F_GA_DFLT_RATES
Projected Minimum Reserve Level	F_FM2000_ANNL
Rate of Reinsurance	F_FM2000_ANNL
Rehabilitation Collections	F_FM2000_MNTH
Total Collections on Defaulted Loans	F_FM2000_MNTH F_FM2000_MNTH_QTR

Table 2: Mapping of Reports to Fact Tables

3 Technical Architecture

3.1 Overview of Technical Architecture

NSLDS, FMS, and PEPS serve as the source systems for data for the FP Data Mart. The initial load of historical information extracted from the FMS system for Guaranty Agencies, and PEPS will perform a routine extract of activity that occurred in these systems from the last time that the extract programs were executed. NSLDS data will be extracted as a 'snapshot' instead of as extracting only the 'incremental changes'. This data will be exported from the source system and placed on the Informatica server using the EAI architecture. Informatica will perform the necessary transformations and then load the data into Oracle, and populate the FP Data Mart. This Data Mart will be accessed by the Microstrategy Intelligence Server to satisfy user requirements that are sent via a web interface. The internal users will access the FP Data Mart directly through FSA's intranet. Below is the diagram of the FP Data Mart architecture. External users will be authenticated prior to allowing access to the FP Data Mart.

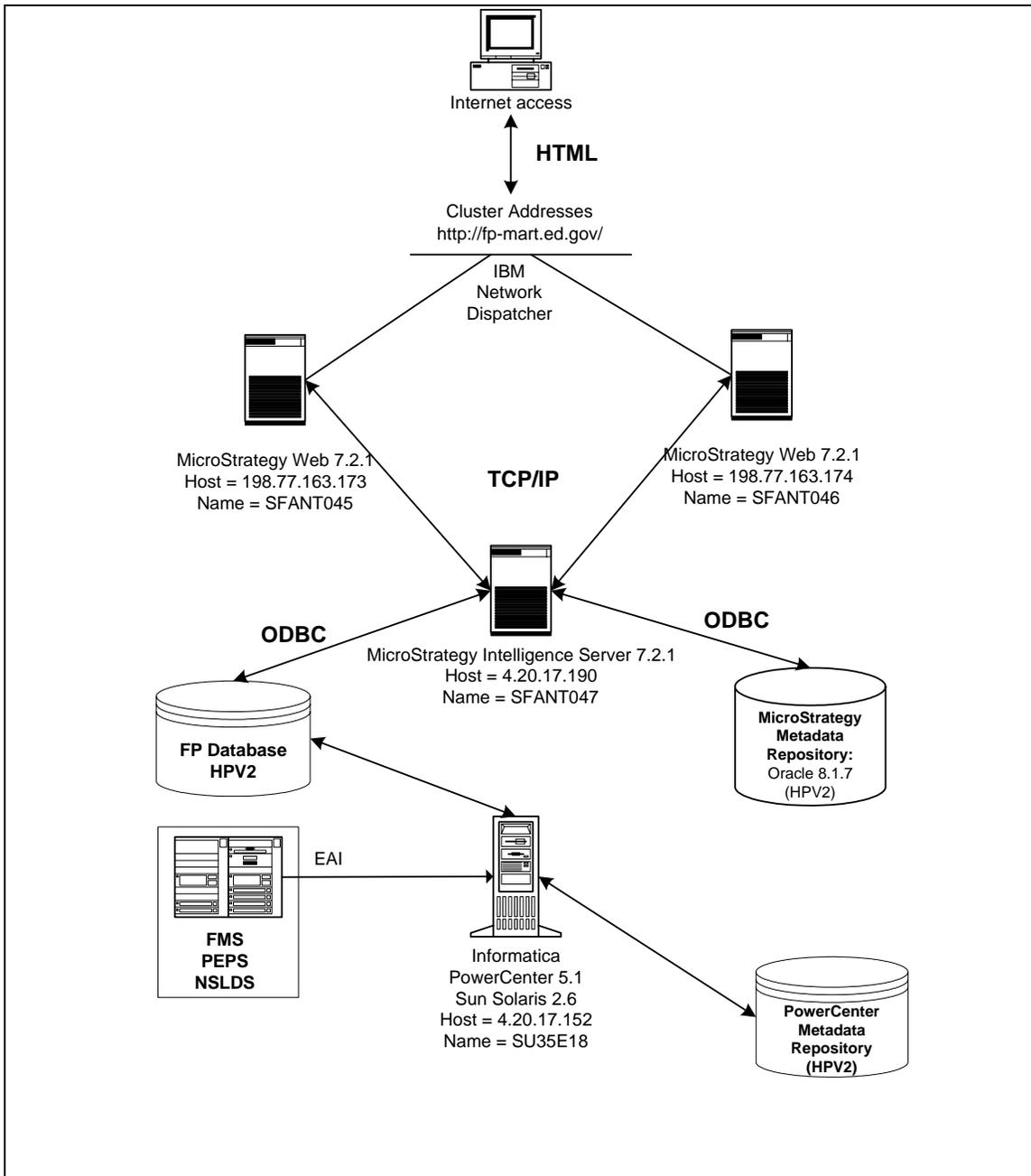


Figure 2: Data Mart Architecture

The FP Data Mart effort utilizes many tools, architectures and processes, including:

- ◆ **User interface via Microstrategy software** - Microstrategy software, a Commercial Off the Shelf (COTS) tool, will provide the front-end user interfaces that allow reporting and On-Line Analytical Processing (OLAP), and the back-end processing including Structured Query Language (SQL) generation, load management, and report scheduling.

- ◆ **Deployment over the web** - Utilizing FSA’s intranet for FP Channel staff and the internet for external organizations, users will be able to log onto Microstrategy software through a web browser. No client-server software installations will be necessary for web-based access to the FP Data Mart.
- ◆ **Use of the Virtual Data Center (VDC)** - The VDC will house the Data Mart development, test and production database and production software servers.
- ◆ **Data acquisition using Informatica** - Informatica, a COTS tool, will use data acquired from the NSLDS, FMS, and PEPS systems. Informatica will read the source data and make the necessary “transformations” to load and populate the FP Data Mart.

3.2 Software

The following table illustrates the software components used to support the FP Data Mart:

Component	Version Information	Installation Tier	Number of Users
Operating System	Sun Solaris 2.6 HP UX 11.0 Microsoft Win2k	Application Server Database Server Application Server	
Compilers	None		
Internet Server	Microsoft IIS		50 concurrent
Database	Oracle 8.1.7	Database	50 concurrent
Application Server	Microsoft Win2k	Application	50 concurrent
Other Application Tools	Informatica 5.1	Application	1
	Microstrategy 7.2.1	Application	25 concurrent

Table 3: Software Architecture

3.3 Hardware

The following table illustrates the hardware components used in the Production environment of the FP Data Mart:

Application	Server Name	IP Address	Manufacturer
Database	HPV2	4.20.15.40	HP
Informatica Server	SU35E18	4.20.17.152	Sun
Microstrategy Intelligence Server	SFANT047	4.20.17.190	Compaq
Microstrategy Web Server	SFANT045	198.77.163.173	Compaq
Microstrategy Web Server	SFANT046	198.77.163.174	Compaq

Table 4: Hardware Architecture (Production)

The following table illustrates the hardware components used in the Development and Test environments of the FP Data Mart:

Application	Server Name	IP Address	Manufacturer
Database	HPN25	4.20.15.15	HP
Informatica Server	SU35E5	4.20.15.135	Sun
Microstrategy Server	SFANT001	4.20.15.244	Compaq

Table 5: Hardware Architecture (Development & Test)

3.4 Middleware

The following figure illustrates how the EAI Architecture will be used to transmit required data from the Release 2 source systems to the Informatica servers at the VDC.

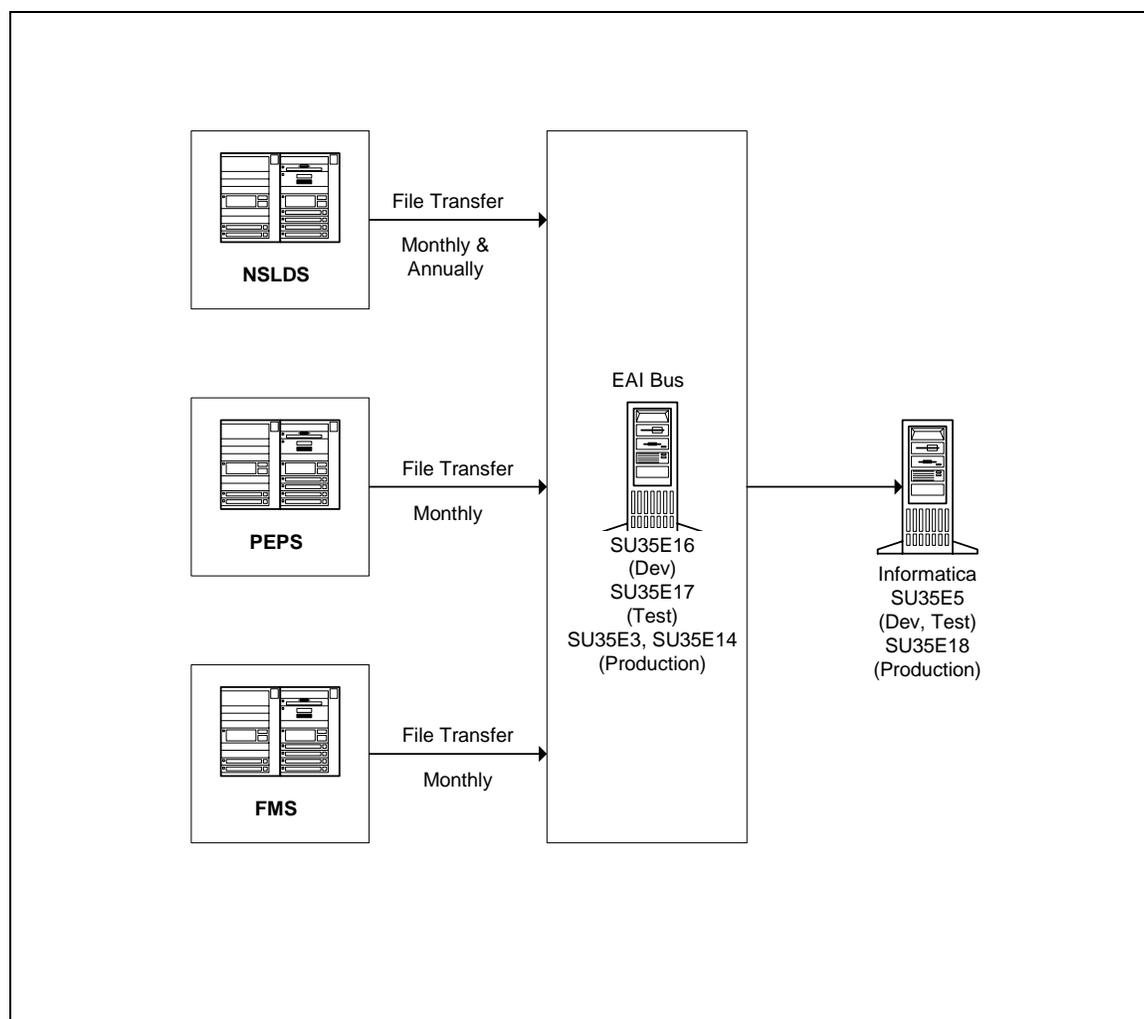


Figure 3: EAI Architecture