FSA Integration Partner

United States Department of Education

Federal Student Aid

Data Strategy Enterprise-Wide Technical Strategies 123.1.10 Data Storage, Management, and Access Strategy

Task Order #123

Version 2.0

October 30, 2003

Executive Summary

Effective data storage, management, and access are key components that support business functionality. Virtually all business processes are either directly or indirectly dependent upon the use of data. FSA identified three primary business objectives to improve data handling across the enterprise:

Rank	Business Objective
1.	Store, exchange, and manage data in a system-independent format that enables access to key data across FSA throughout all phases of the lifecycle
2.	Provide data access to varied resource needs, in the formats necessary to provide meaningful business information and provide data mining and analytics tools that allow simple access to data and facilitate trending, forecasting and the necessary information for business decisions.
3.	Improve the timeliness and efficiency of data feeds to accommodate varied audience data needs.

Data Storage, Management, and Access Business Objectives

In today's environment, FSA stores and manages data in several disparate systems. Similar data is often duplicated many times across the enterprise. Student demographic information, for example, is stored in the CPS, COD, and NSLDS systems. At various times during the student aid lifecycle, any one of these systems may update student information, but does not always replicate the change throughout the enterprise. Storing similar or overlapping information in multiple places often leads to systems presenting different versions of the truth.

Another consequence of this data environment is that FSA does not have visibility to key information across the student aid lifecycle. Data is stored in multiple systems depending on the stage of the process. Users must access multiple independent systems in order to gather a complete picture of a student or trading partner.

Such limitations were drivers for developing a set of options for the long-term data storage architecture. FSA business owners reviewed and updated these options in a series of retreats and working sessions. The options represented a wide range of solutions with varying levels of data consolidation. The "Common Access" option was selected as the target FSA end-state.

Recommendation

The Data Storage, Management, and Access Strategy recommendation supports the long-term vision through the following grouped recommendations:

- Store: Develop a Common Data Architecture
  - Build an enterprise data warehouse to store historical information from all phases of the FSA lifecycle.
  - Store shared data to support all front-end business processes.

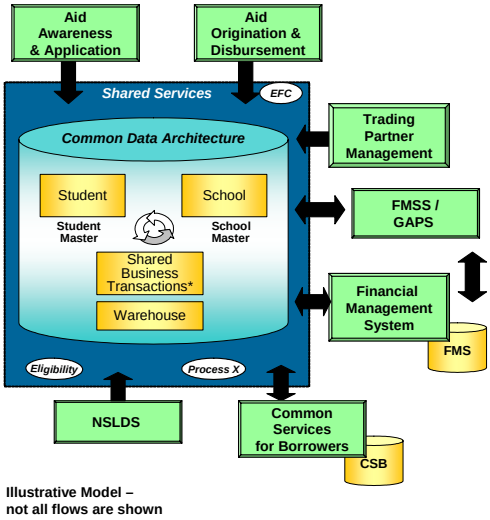
- Access: Phase-in Uniform Business Intelligence Capabilities Across FSA Organization
  - Create a single point of access for FSA data.

- Provide tailored access to a broad range of FSA user groups (e.g. reports, ad hoc queries, executive dashboard).
- Expand centralized web-based access to historical FSA data.
- Create process-specific data marts sourced from the enterprise data warehouse.

- Integrate: Use Data Storage Solution to Simplify Future Initiatives
  - Enable system and user access to the Common Data Architecture.
  - Leverage internal integration capabilities to support business service requests.
  - Consolidate data definitions (metadata) from multiple systems to provide business context and common definitions to users.
  - Gather feeds from both front- and back-end systems to compile cross-lifecycle information.

- Enable: Develop an Enterprise Data Warehouse Framework
  - Implement robust framework to address all facets of warehouse
  - Build scalable solution to accommodate the increasing scope of shared data and a rising number of students.
  - Long-term solution must accommodate a growing user base as functionality is rolled-out to the broader FSA community.

The target-state FSA data architecture represents a move toward centralized data storage and business process alignment. This vision represents a shift from system-focused to business process-focused functionality. This target-state FSA architecture centers on a Common Data Architecture that will provide a consistent set of data, available to FSA business owners and systems. Data will be stored once in a central location rather than spread across many systems.

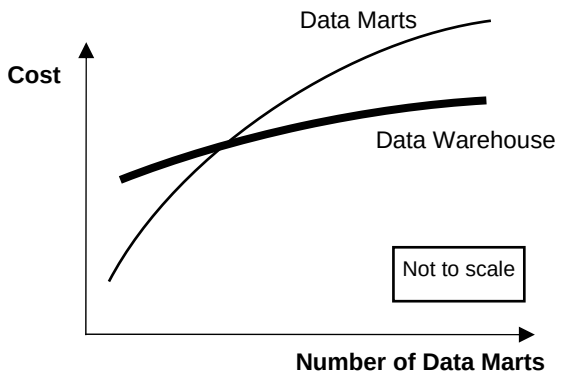


Long-Term Vision

Industry trends show organizations are turning to data warehousing initiatives for cost savings and improved use of data. A recent analysis by Gartner finds:

"Several business benefits — including cost reduction, increased information accuracy, information timeliness and user productivity — are enhanced though data warehouse and business intelligence deployments that support flexibility and durability."[1]

Enterprises are seeking to reduce data storage costs and provide wide-visibility reporting capabilities by consolidating overlapping data storage environments and by combining data from disparate business functions or divisions. Limited-scope data marts can provide initial access to data; however, a large number of data marts without a central data warehouse can adversely affect operational efficiencies. The figure below shows the relative costs of a data warehouse compared to multiple data marts:



Relative Cost of Data Marts[2]

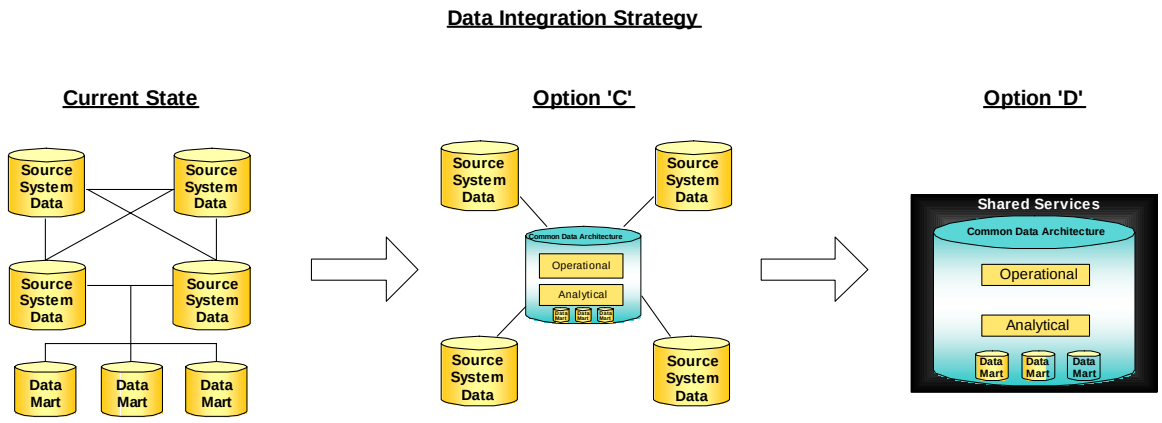
A subject-area data mart, due to its limited scope, is typically easier to implement than an enterprise data warehouse. With the creation of additional marts, however, maintainability and scalability issues force the overall data mart costs higher. A comparison of costs of building a data mart versus a data warehouse shows that the total cost of building the equivalent number of data marts may exceed the cost of building the original data warehouse.

Merely storing data serves little purpose without providing efficient means of accessing the data. In order to provide business owners the visibility they need to the data, FSA should deploy an enterprise business intelligence (BI) solution on top of the Common Data Architecture. Enterprise-wide access to data will include access methods customized for each user's requirements. This could range from simple reporting to high-level, aggregate executive dashboards. FSA should manage appropriate user access to sensitive data in conjunction with the security framework and access management initiatives.

[1] © Copyright Gartner, Inc. Source: "Management Update: Total Value of Opportunity Determination Is Key for Data Warehouse Planning?" September 17, 2003.
[2] © Accenture, Source: "BI Strategy Assessment" January 2003

FSA should leverage the consolidated data environment to support cross-lifecycle activities. Information about stored FSA data (i.e. metadata), would help provide business context to users of the data. In the present state, different systems independently generate this "data about data". FSA should consolidate this information to enable users to make the most use of the vast amounts of data.

How do we get there?
A gradual evolution to the end-state solution will allow FSA to enable pieces of individual functionality, realize interim benefits, and minimize business process impacts. In today's environment, data is stored independently on many systems. FSA should consolidate much of this data storage into a centralized data storage environment through an evolutionary approach:



Front-End Data Storage Changes Over Time

Front-end systems, those related to the application, origination, and disbursement of student aid, would evolve to the Consolidated Data Store over time. In the near-term solution, the individual systems retain their own data repositories and business processes but post key information, such as student and school demographic data, to a shared system. In this solution, the Common Data Architecture has stewardship of the student and school data, which other FSA systems use as their master record. This reduces data errors, as there is only one repository for the true version of the data. As the solution evolves, the shared storage environment grows while the system-specific data stores shrink. In the target-state, all FSA front-end systems use the Common Data Architecture without permanently storing any data at the system-level.

As FSA data consolidation progresses, managing access to the data becomes easier. A single enterprise-wide solution is used to provide access to the Common Data Architecture to business owners throughout the organization. Individual access to the data should be tailored so that each group of users receives the data in an appropriate format.

The successful evolution from the current-state to a consolidated data storage environment requires consideration of several factors. Scalability of both the user base and the volume of data is critical. Hardware and software decisions need to be made with the understanding that they will support a growing environment. Data storage and access platforms have different scalability approaches. These need to be closely examined as part of detailed planning efforts.

Benefits
When fully implemented, the Data Storage, Management, and Access Strategy will provide benefits including:

- Improved Data Visibility
  - Deliver analytical capabilities (detailed reporting to executive dashboards) for business owners to gain visibility across the enterprise.
  - Reduce data gathering effort by tailoring access methods to individual user needs.
  - Improve cross-organizational insight to help understand the state of business and enable quick response to issues that could place FSA on the GAO high-risk list.

- Reduced Maintenance Costs
  - Eliminate redundant data storage; all front-end business processes will act upon the same set of data.
  - Reduce data synchronization processes between systems.

- Simplify Integration Within FSA
  - Simplify new projects by leveraging a consolidated source of data, rather than many independent systems.
  - Leverage security & access management initiatives to provide appropriate user access to data.

- Improved Quality and Usability of FSA Data
  - Reduce data quality issues by storing data once, in a consolidated environment.
  - Optimize performance by separating operational and analytical functions into distinct storage environments.

IMPROVED DATA VISIBILITY
Giving FSA business owners access to an enterprise-wide source of data for reporting purposes will provide a clearer understanding of the state of the business. While the NSLDS and data marts currently provide data for these analytical needs, access is limited in two ways. The first of which is that the underlying data typically focuses on a certain aspect of the aid lifecycle and does not provide a comprehensive picture over time. The target-state addresses this limitation by using components of the Common Data Architecture as the foundation for reporting capabilities across FSA. A single source will provide data from a wide range of FSA business processes instead of multiple system- or process-specific data stores. The second limitation results from the data access capabilities not being rolled-out to the entire enterprise. The target-

state solution will deploy the same business intelligence (BI) toolset across the organization. This will leverage a BI toolset's flexibility to provide access to FSA data in a format tailored to the individual user's needs.

REDUCED MAINTENANCE COSTS

In today's environment FSA maintains data storage on several front-end systems. As data storage shifts to the Common Data Architecture, the distributed storage systems will become smaller to the point of being completely transferred to centralized storage. As this solution progresses, more and more of the maintenance and operations effort will be consolidated into the shared environment. Standardization of architecture platforms, support, and development should allow for a reduction of maintenance effort.

SIMPLIFY INTEGRATION WITHIN FSA

The proposed data storage solution will provide a stepping-stone to initiatives that previously would have been challenged to access cross-lifecycle data. The long-term goal of centralized data storage will enable many initiatives to reduce development effort by providing a single source for data. Without the consolidation piece, other integration efforts would be required to tie to several FSA systems and processes to gather the data needed for each business process.

For example, prior to the realization of the target state, the ability to provide a student and school view would be dependent upon the ability to obtain and aggregate information from multiple FSA systems. Once the target state is fully implemented, other systems will not need to aggregate data from many sources; the data warehouse will store aggregated information in a common location. The consolidation of data storage will support other projects by allowing them to build a single connection to enterprise data rather than creating multiple interfaces.

IMPROVED QUALITY AND USABILITY OF FSA DATA

By leveraging a common set of core components and by storing data a single time, FSA can standardize data definitions and reduce many common data quality concerns. Enabling a single source of data will eliminate many of the present-day problems with separate systems presenting different data for the same piece of information.

Separating operational or near-real-time processes from analytical data stores will allow FSA to optimize each environment for its specific use. The operational environment will be designed to quickly process many transactions or updates, while the analytical environment can be optimized for query performance. Optimizing the systems for their core functionality will allow FSA to design for a specific purpose and increase the performance of each environment.

Next Steps

The successful evolution from the current-state to a consolidated data storage environment requires a detailed planning effort that considers several factors including: the scalability of both the user base and the volume of data, hardware and software decisions to be made with the understanding they will support a growing environment, and scalability approaches for data storage and access platforms. Without a detailed action plan that addresses these

considerations, FSA's evolution to the target state and realization of the benefits of a Consolidated Data Architecture will be jeopardized.

The Data Storage, Management, and Access Strategy will be combined with strategies from the other four technical areas to represent a comprehensive Technical Strategy that is in alignment with the overall FSA Data Strategy effort. The collective picture and the required implementation steps will be outlined in the Technology Vision and Strategic Plan (123.1.12), Data Framework Technical Specification (123.1.4), and Quality Assurance Strategy and Implementation Plan (123.1.5). These documents will serve as the enablers of the high-level Data Strategy business objectives.

In preparation for the delivery of the Technology Vision and Strategic Plan (123.1.12), working sessions will continue to be conducted with key FSA stakeholders through October as part of the effort to develop the Technology Vision and Strategic Plan, which will be delivered in mid-November.

Amendment History

DATE	SECTION/ PAGE	DESCRIPTION	REQUESTED BY	MADE BY
10/14/2003	All	Document submitted for FSA-wide review.	N/A	Kyle Michl
10/30/2003	Sections 1.5, 2.2, 5.1, 5.5, and 6..1	Changes incorporated based on feedback provided by Denise Hill.	Denise Hill	Kyle Michl

Table of Contents

1 INTRODUCTION 4

1.1 PURPOSE 4
1.2 BACKGROUND 4
1.3 SCOPE 4
1.4 ASSUMPTIONS 4
1.5 BUSINESS OBJECTIVES AND GAPS 4
1.6 KEY DECISION POINTS 4
1.7 SUBJECT MATTER EXPERT UTILIZATION 4

2 WHAT IS THE OVERALL VISION FOR CONSOLIDATED DATA SERVICES? 4

2.1 EVOLUTION OF DATA STORAGE 4
2.2 THE TARGET STATE 4

3 HOW WILL THE DATA STORAGE SOLUTION SUPPORT AND IMPROVE BUSINESS CAPABILITIES? 4

3.1 DATA ACCESS CONCEPTS 4
3.2 HOW CAN DATA ACCESS SUPPORT BUSINESS CAPABILITIES? 4
3.3 IMPACTS OF A DATA WAREHOUSE SOLUTION 4

4 WHAT ARE THE INTEGRATION CONSIDERATIONS FOR THE COMMON DATA ARCHITECTURE? 4

4.1 COLLECTING AND PRESENTING DATA 4
4.2 PROCESSING AND MOVING DATA 4
4.3 DESCRIBING DATA 4

5 HOW WILL THE DATA BE STORED? 4

5.1 DATA ARCHITECTURE CONSIDERATIONS 4
5.2 INTEGRATION STEPS 4
5.3 PERFORMANCE 4
5.4 STEWARDSHIP OF DATA 4
5.5 TECHNOLOGY SUPPORT FOR DATA ACCESS AND STORAGE 4

6 RECOMMENDATION SUMMARY 4

6.1 RECOMMENDATION DETAILS 4
6.2 BRIDGING OF GAPS 4
6.3 ALIGNMENT WITH BUSINESS OBJECTIVES 4

APPENDIX A: BUSINESS OBJECTIVE ACCOMMODATION CRITERIA 4

APPENDIX B: MEETING MINUTES – WORKING SESSION #1 4

Tables and Figures

Table 1.1 - Data Storage, Management and Access Current State Gap Analysis 4
Table 6.1 - Recommendation Gap Bridging 4
Table 6.2 - Recommendation Business Objective Fulfillment 4

Figure 2.1 - Evolution of Common Data Architecture 4
Figure 2.2 - Data Integration Strategy 4
Figure 3.1 - Business Intelligence Framework 4
Figure 4.1 - Web Usage and Web Services Integration Example 4
Figure 4.2 - Integrated Services and Data Storage Example 4
Figure 4.3 - ETL Example 4
Figure 4.4 - External Information Access and Data Storage 4
Figure 6.1 - Data Storage, Management and Access Phased Approach 4

1 Introduction

1.1 Purpose

The purpose of the Data Storage, Management, and Access Strategy is to outline options to help FSA reach its data integration vision. This document will expand on the current state and business objectives related to data storage, management, and access that were captured in the Technical Strategies Statement of Strategic Focus (123.1.6) by clearly documenting options that can be used to fill gaps that exist between the current and target states.

Right-time access to data is the foundation from which FSA can build out its financial aid services and provide valuable customer service to students and trading partners. This strategy reviews how FSA can establish an efficient and versatile data architecture that meets the enterprise's business needs for storing and accessing data. It also provides insight into how FSA can combine the implementation of technology with business processes to increase business decision efficiency. The Technical Strategies review of data storage, management, and access supplements the work in progress by other Data Strategy efforts and takes into consideration ongoing projects at FSA. Specifically, this document supports the future direction of FSA's data exchange vision as captured by the Data Framework team.

Furthermore, this document addresses FSA Performance Plan Action Items 16.5 (Develop integrated Data Warehouse and Data Mart strategy), 16.5.1 (Develop Data Warehouse and Data Mart Vision), 16.5.2 (Develop Data Warehouse and Data Mart Framework), and 16.5.4 (Develop a strategy for enterprise-wide data sharing and distribution [data storage, management, and access rights]). The document also includes options assessments of solutions that may help FSA align its key technology areas with its overall business objectives. This document is part of a series of five technical strategy documents. This Data Storage, Management and Access Strategy will be combined with strategies from the other four technical areas to represent a comprehensive Technical Strategy that is in alignment with the overall FSA Data Strategy effort. The collective picture and the required implementation steps will be outlined in the Technology Vision and Strategic Plan (123.1.12), Data Framework Technical Specification (123.1.4) and Quality Assurance Strategy and Implementation Plan (123.1.5). These documents will serve as the enablers of the high-level Data Strategy business objectives.

1.2 Background

The Department of Education's Federal Student Aid (FSA) organization is seeking to deliver overall improvements in the areas of data quality and data consistency. The goal of implementing improvements to the flow and consistency of data exchange is to ensure that FSA complies with regulations set forth by oversight organizations in support of the program-wide goals, which include maintaining a clean audit and ensuring exclusion from the General Accounting Office (GAO) high-risk list.

The purpose of the FSA Enterprise-Wide Data Strategy is to define the FSA enterprise data vision and strategy for how it will combine the tools, techniques and processes, documented in

the FSA Data Strategy Framework, to handle its enterprise data needs. The effort to store, manage, and provide access to data at FSA is one of five technical strategies that, when combined with the strategies for Data Framework, XML Framework, Common Identifiers, and Enrollment and Access Management comprise the overall FSA Enterprise-wide Data Strategy. The technical strategies in addition to Data Storage, Management, and Access include Internal Data Exchange, Web Usage (Portals), External Information Access (FSA Gateway), and Web Services. Each of these strategies are addressed in separate documents, but were considered during the development of the overall Data Storage, Management and Access Strategy.

In order to create the overall Data Storage, Management and Access Strategy for FSA, the current state of data storage was compiled and validated with key business owners. The current state served as a basis for a series of business objective gathering sessions. Detailed business objectives were gathered with respect to the five technology areas from FSA business owners across the student aid lifecycle.

Approximately, two hundred raw business objectives were gathered for each of the technical strategies key areas. These business objectives were then refined, consolidated, and prioritized by FSA business owners in a series of consensus meetings. The business objectives and associated gaps between the objectives and the current state of technology at FSA were gathered and presented through the Technical Strategies Statement of Strategic Focus (123.1.6).

This document details how improvements in data storage, management, and access techniques can help FSA achieve the business objectives that were set forth in the Technical Strategies Statement of Strategic Focus.

1.3 Scope

The goal of this strategy document is to provide insight to the considerations that must be addressed to improve the method for storing and accessing data across the enterprise. The strategies associated with each of the five technical areas will be combined into the Technology Vision and Strategic Plan (123.1.12) to present an overall technical strategy along with a road map for achievement. The key items that this document will present include:

- An overview of the Data Storage, Management and Access current state with respect to business objectives.
- A detailed options assessment to outline the future state of enabling Data warehousing/data mart technologies.
- A review of conceptual direction set out by business owners and an assessment of technical considerations necessary to realize that conceptual vision.
- Recommended technical options to help FSA best achieve the business objectives.

1.4 Assumptions

The following assumptions have been factored into the scope for the Data Storage, Management and Access Strategy:

- Although there is tangible value that can be achieved in the short term, FSA's Data Storage, Management, and Access vision includes characteristics that may need a three to five year timeframe to implement in order to fully achieve the organization's business objectives.
- Since the time of the original business objective gathering sessions, additional projects have been identified as having an impact on the Technical Strategies and FSA's future state. As appropriate, these efforts and other potentially relevant efforts will be assessed for impact during subsequent Technical Strategies deliverables.
- Considerations have been made that account for the future FSA direction of Internal Data Exchange, Web Usage, External Information Access, Web Usage and Web services.

1.5 Business Objectives and Gaps

FSA manages a variety of storage techniques to collect and disburse financial aid information. There are many operational systems, some of which perform data warehousing/data mart functions. There is also a set of data marts that are populated from source systems to enable analytical reporting. In order to improve data storage and access, a set of consolidated and prioritized business objectives from the consensus meeting was captured in the Technical Strategies Statement of Strategic Focus (123.1.6). These objectives then served as the primary input for the overall Data Integration Vision.

FSA's vision for the future of data storage and access is driven by the need for consolidated, system-independent data access and improved analytical capability. This strategy focuses its technical options assessment on the ability for the recommended solutions to fulfill the gaps and provide FSA with a direction to reach its business objectives. Based upon the feedback obtained from key FSA business owners, the following table outlines the business objectives and gaps that must be filled to reach the target state. Explanations of the assessment criteria can be found in Appendix A: Business Objective Accommodation Criteria.

Rank	Business Objective	Percent Realized	Gaps
1.	Provide data access to varied resource needs, in the formats necessary to provide meaningful business information and provide data mining and analytics tools that allow simple access to data and facilitate trending, forecasting and the necessary information for business decisions		• No customer end-to-end visibility across the Financial Aid Lifecycle. • No consistent use of business intelligence tools to access data. • Minimal user access to create queries and mine data as needed.
2.	Improve the timeliness and efficiency of data feeds to accommodate varied audience data needs.		• Data feeds occur at varied frequencies and are not in the required order. • Un-reconciled data is often transferred.

Rank	Business Objective	Percent Realized	Gaps
3.	Store, exchange, and manage data in a system-independent format that enables access to key data across FSA throughout all phases of the lifecycle.		• No enterprise-wide view/store of data exists. • Data mining capabilities are limited. • FSA control of data is impacted by varied third party systems and owners. • Current data storage architecture will not fully support enterprise data needs outlined in target-state.

Table 1.1 - Data Storage, Management and Access Current State Gap Analysis

1.6 Key Decision Points

Based upon the business objectives and gaps that were identified in the previous section, a series of key decision points were established and refined by key cross lifecycle business owners. The purpose of these decision points is to provide guidance in developing options and proposing a solution:

- **What is the FSA vision for data consolidation?**

  This strategy documents FSA's overall vision for data storage and its ability to provide access to data services. It reviews the outcomes from the Technical Strategy working sessions in which key FSA stakeholders identified a data integration vision for the future state of data storage, management, and access. The data storage vision works in conjunction with the Data Framework team's recommendations for data exchange across the enterprise.

- **How will systems and people access data?**

  This strategy reviews this question in order to assess how FSA will provide the right information to the right person in the right format at the right time. It is important that data not only be stored and managed in an efficient manner but users also need the ability to access the data to accomplish their business needs.

- **What are the integration points to other FSA entities?**

  The concept of data storage, management and access has impacts across the enterprise beyond just the configuration of the architecture. This strategy reviews areas across the enterprise that are impacted by data storage and access techniques.

- **How will the data be stored?**

  One of the highest priorities for FSA is to improve data quality across the enterprise. In order to minimize data inconsistencies, FSA must review the data storage architecture to ensure that information is efficiently collected and stored. The data storage vision takes

into consideration the potential for a reengineering of FSA's systems and processes as well as the need to continue to provide timely and accurate access to information.

1.7 Subject Matter Expert Utilization

The following Subject Matter Experts (SMEs) were consulted during the creation of the External Information Access Strategy

1. Shiping Wang, Data Warehouse Expert – Accenture Global Architecture & Core Technologies Group
2. Mohit Sahgal, Data Warehouse Expert – Accenture Solutions Engineering
3. Jonathan Hill, Enterprise Integration Expert – Accenture Global Business Solutions

2 What is the Overall Vision for Consolidated Data Services?

Improving data quality, consistency, access, and movement were the underlying drivers for the Data Strategy effort. Data is a key component of successful business execution. As such, changes to the storage and access of data have a big impact on existing business processes and require a high level of effort to change. The overall vision of data must look at the various types of data. For example:

- Operational data: Data that is acted upon multiple times a day by FSA systems to complete business functions
- Analytical data: Data that is not actively updated by systems, but is required to perform historical analysis, trending, or historical "checks."
- Metadata: Data that describes the characteristics of a resource associated with electronic systems. This can be of a business (e.g. data dictionary, calculations, process steps) or technical (e.g. hardware/software information, process statistics, etc.) nature. Metadata is often described as "data about data".

The FSA vision for data must take into consideration these types of data as well as address these impacts:

- Business processes need to change to support new storage options.
- System ownership of data will change.
- Data storage infrastructure needs to support both shared transactional access and common access to data warehouse.
- Data architectures must scale and grow as the enterprise changes.
- Management of complex data relationships in a single location.

2.1 Evolution of Data Storage

Evolution from a system-oriented environment to business process-based execution implies not only change in the systems and business processes, but in the management of data as well. Complimentary to consolidating business processes is the reduction of redundant data and sharing of common data. In response to the call for effective data management, Gartner identifies that "Enterprises are trying to simplify the IT environment and data management platforms, exploring ways to reduce the complexity of multiple DBMS products. With each reduction in DBMSs comes migration and possibly data consolidation."[3]

There is a trade off between complexity of integration and the complexity of operations management. The Data Strategy core team explored this concept at a retreat on Thursday, August 21st. The group reviewed a series of options (A-D) that showed an evolving progress from system-owned data, to centralized access, to a common data store by business process. The group confirmed that data storage is a foundational piece to enable their broader data

[3] © Copyright Gartner, Inc. Source: "Database Management and Administration Key Issues." February 4, 2003.

integration strategy that includes business processes and shared services across the enterprise. The Data Framework Specification (123.1.4) reviews these options in the context of the overall future state for the FSA business architecture, whereas this strategy reviews the evolutionary process for the data storage component of FSA's strategic vision as it relates to these options. The following diagram depicts the evolutionary process from the current state of data storage:

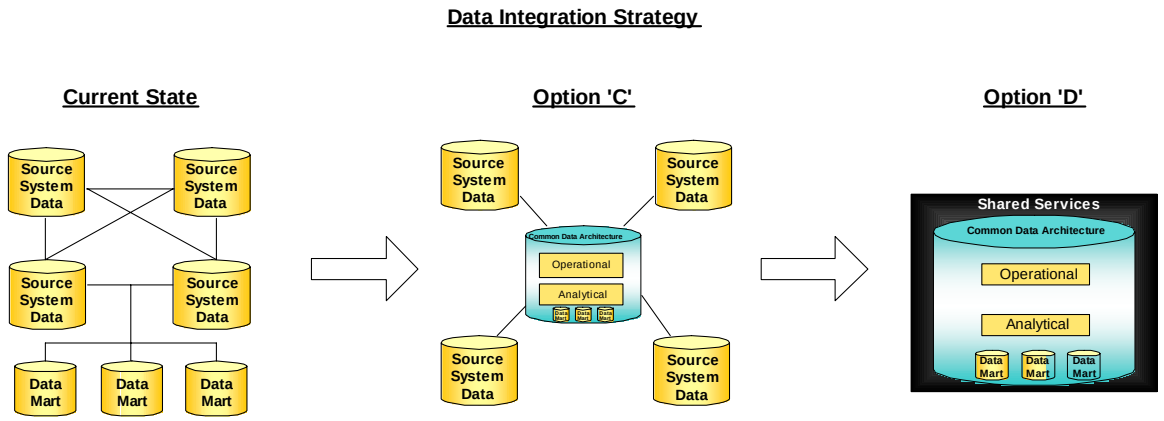


Figure 2.1 - Evolution of Common Data Architecture

2.2 *The Target State*

FSA business owners and the Data Framework team have identified a long-term data strategy with impacts to the storage, access, and management of data. During the retreat, FSA concluded that for its long-term vision, true centralized storage of and access to common data would be the ideal state (Option "D") to facilitate a reduction in redundant data and enable common business function access. This option will enable service-oriented access to operational data, with real time access to data as needed by each business process. This architecture will also eliminate the many concerns around data consolidation between multiple front-end systems, as they will share a common source of operational data. Thus, the data storage solution will be comprised of both operational and analytical data; the distinction being that operational data is frequently updated and is used to support core business processes. Analytical data structured differently and is updated less frequently in order to provide a more stable environment for historical data access. Key data, such as student and school information, will reside in both places, but the historical view of both will be in the Analytical Data Warehouse component of the CDA. The physical architecture of the storage solution is a design issue out of scope for this deliverable. The details pertaining what information is in each storage component need to be defined as part of a design effort.

Iterations and permutations of options A through C will be required before Option D can be achieved. The technical strategy for achieving the end-state needs to account for this continuum of evolving solutions. This iterative process takes time and careful planning to mitigate impacts and ensure business execution remains efficient during the evolution. Furthermore, Option C was highlighted as an interim step that would enable a controlled migration towards the future state. The technical details of the implementation must be reviewed in depth prior to implementation, however, prior to then, a firm understanding of the design and implementation considerations can enable intelligent decisions upfront. Section 5 discusses in detail the considerations that should be reviewed prior to moving forward with iterative integration steps. The following diagram illustrates the evolutionary continuum:

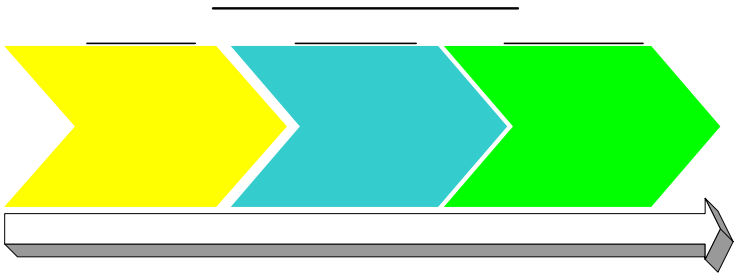


Figure 2.2 - Data Integration Strategy

3 How will the Data Storage Solution Support and Improve Business Capabilities?

The Common Data Architecture, contains operational (student master, school master, and business transactions) and analytical (data warehouse, data marts) components. The analytical portion of the architecture will provide individuals with access to data about the state of the business in their environment. The operational components of the data architecture will provide a consolidated data environment that will simplify business process and act as an enabler for providing consistent business services.

The high-level vision for the Common Data Architecture addresses the relationships between FSA systems, and how they will share and store data. The mere storage of data, however, does not necessarily create business value. A critical piece of the Data Strategy is how this data will be put to use -- specifically, how will systems and people be able to *access* this data and use it to enable business functionality?

3.1 Data Access Concepts

The concept of business intelligence (BI) refers to the ability to obtain meaningful information from disparate data across the enterprise that can provide business insight to an organization. There are many products that must work together to provide these capabilities. It is important to understand the various components that make this data access meaningful. The Business Intelligence Framework, as illustrated in Figure 3.1, provides a summary of key concepts relating to data warehouse and business intelligence implementations. These concepts support the gamut of business challenges including day-to-day operational decisions, decisions based on changing market conditions, and strategic, long-range decisions. The figure shows business intelligence capability increasing from operational reporting to strategic real-time intelligence systems as you move from left to right on the continuum.

This framework provides a useful model of the possible capabilities enabled as FSA progresses from the as-is state, to a centralized, common data architecture.

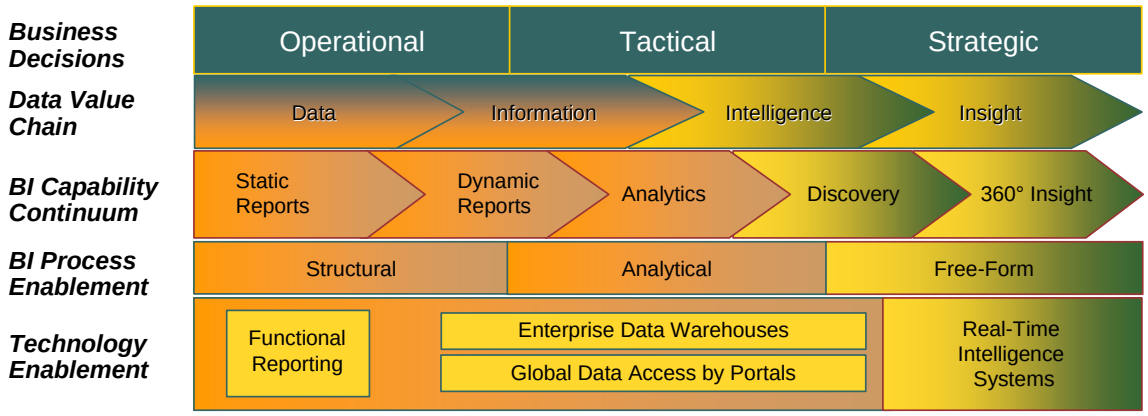


Figure 3.1 - Business Intelligence Framework[4]

As FSA evolves its business intelligence capabilities, it can expect greater value from the data warehouse. Early implementations would fall under the Operational portion of the framework, which would have static and dynamic reporting capabilities. These reports would give FSA business owners a clearer vision of the underlying state of business. Structured data requirements will provide the basis for these reporting procedures. FSA currently has this capability in a limited sense: analytical reporting access exists for specific business capabilities in the custom data marts. The three data marts in today's environment enable end-user access to specific functional groups of data. A problem with the current-state is that it is very difficult to accommodate new data requirements. For instance, the interfaces to the source systems will need to be modified if a piece of data is not currently included in the data mart, the interfaces to the source systems have to be modified. Funding issues related to changing the source system feeds can delay the implementation of new data. As the implementation of the Common Data Architecture progresses, these analytical support capabilities will become more and more integrated and provide access to a more enterprise-wide base of analytical data.

As FSA moves to the Strategic data focus shown on the framework, more advanced options for data access are enabled. Business Intelligence will begin to offer more tangible business value, as business owners are able to gain new insight into FSA. New capabilities such as uncovering previously unknown relationships (discovery), and providing 360-degree insight, will better enable end-users to discover new information without being limited to pre-defined, structured queries. 360-degree insight is the confluence of many business intelligence efforts that enables an organization to have full end-to-end understanding of their business processes. It could enable the organization to better understand the relationships between FSA business entities (borrowers, trading partners, etc.), and predict how each business decision may affect them.

Business Intelligence Framework Highlights

- The BI framework supports many types of business decisions, from operational to strategic.

[4] © Accenture "Business Intelligence Framework & Key Concepts", October 2002

- Each movement to the right in the data value chain adds value by supporting more advanced decision-making.
- The organizational decision making capability increases as BI capability becomes more sophisticated and more readily accessible.
- Advanced business intelligence capabilities enable users a less-restrictive, "free-form" environment in their search for data. Compared to the highly-structured processes at the left of Figure 3.1, free-form BI allows greater flexibility with less restrictions as to how data can be accessed.
- The technology enablement piece is comprised of the underlying analytical storage environment and possibly business process management components.

As a business intelligence solution progresses from the simplistic to the sophisticated, greater capabilities are enabled. For realization of value, the enterprise must be able to use the data effectively. Additionally, there must be a business need for the investment required to get the most out of the components of Business Intelligence. There are a number of methods available to access information in a data warehouse or a data mart:

- Pre-defined queries are simply queries against the database that are written in advance. These queries can shield the end-user from any deep knowledge about the structure of the data warehouse, but they do not offer much flexibility in the use or content of data provided. A separate data warehouse power-user or support staff is required to develop the query. Many end-users can then share these queries without having a detailed, technical understanding of the data structure.
- Static reports are based off pre-defined queries, but are generally in a more presentation-ready format. Static reports are often generated automatically and distributed to a less-technical group of users.
- Ad hoc queries allow users the greatest flexibility to access the data. Ad hoc query is a term for those queries that are created on demand. There is no pre-definition; users are allowed to create their requirements from scratch. This method of database access is generally reserved for a group of "power users" who have had specialized training. Ad hoc queries are limited because there is little control over what impact they have on a database. Several power users simultaneously executing very complex queries could impact the performance of the overall solution. If the ad hoc query access method is used by those without a detailed understanding of the data, inaccurate information can be obtained.
- Drill-down reporting provides detailed access to data in a user-friendly manner. More comprehensive than a static report, a drill-down report allows users to investigate the data more in-depth. An example of a drill-down is a report that shows a monthly total; the user could click on the total to see the individual components which, added together, comprise the total.
- Cube analysis is an OLAP data warehouse/data mart access method allowing users to view dimensional data in a logical manner. End users are allowed to navigate the data by viewing facts related to each dimension.

- Executive dashboards provide a high-level presentation of key business factors. Customization of a dashboard can allow executives to decide what information is important to them and view it in a clear and concise manner.
- Data mining is a technique for analyzing data to uncover patterns to help predict future behavior. Data mining is typically reserved for a highly trained group of users working to improve business processes or identify new trends.
- Event-based triggering allows a database to proactively push data or an alert to another system or to an end user based on changes to the data. The data warehouse sending a warning notification when the amount of borrowers in default crosses a threshold is an example of event-based triggering.

The relative strengths and weaknesses of each of these methods should be taken into account when deciding what access methods should be made available to each group of users.

3.2 How Can Data Access Support Business Capabilities?

Data is of little value without allowing proper FSA resources to *use* the data. In the end-state vision, front-end FSA systems will use the operational components of the common data architecture to support real-time transaction processing. The Student and School Masters, and the Business Transactions make up the operational component of the Common Data Architecture. These data stores allow FSA to begin to consolidate the multiple front-end system data. Many of today's concerns, such as which system has the most current picture of the data and the increasing complexity of cross-system data feeds will be more easily accommodated in the future state model. If the data is only stored in one place, the questions regarding out-of-sync conditions become moot. Data consolidation will also reduce concerns around the management of several data feeds between systems. These systems will all operate directly against the operational components of the Common Data Architecture.

Generation of the Individual Student Information Record (ISIR) is one business process that could benefit from the integration of storage and management of operational data. The Central Processing System (CPS) has the primary responsibility for the generation and distribution of the ISIR. The ISIR is initially sent to all of the schools to which the student has applied. A number of factors can change aid package options, including the school making additional grants, or a change in the employment status of the student's family. These changes will affect the ISIR and can be communicated back to FSA either through CPS or through the Common Origination and Disbursement system (COD). Depending on which FSA channel was informed of the change, the updated ISIR information could be contained in either the CPS or COD system. The consolidated operational data storage of the target vision could help fix this out-of-sync condition because neither CPS nor COD would contain proprietary views of the application. Instead, all FSA front-end systems would use the same set of data for their processes; regardless of the point of entry, the data would be updated in the integrated operational components of the Common Data Architecture. Any subsequent generation and/or distribution of an ISIR for that student would be based on the up-to-date information in the centralized data store. In the interim state, before the full implementation of the Common Data

Architecture, the consolidated operational store could be able to provide a "best answer" to questions related to ISIR status. Although the individual business processes would not be using the shared data for operational support, the operational components of the Common Data Architecture would still receive status updates from the various systems and could provide status across the enterprise.

Possible Benefits:

- "One view of the truth" – consolidated data reduces or eliminates chance of conflicting enrollment status
- Propagation of updates – Changes to student information are available to all entities; no more containment of data in which only one system is aware of recent updates
- Reduced data volume – The enrollment data is only stored once, not replicated in several systems

Analytical data access will allow users and systems to browse or query historical FSA data. Creating a centralized data warehouse will enable a consistent picture of FSA data without having to interact with many loosely connected systems. Both front-end data from the operational components as well as back-end data on the servicing of loans will populate the data warehouse. The data warehouse will allow end-users to access broad categories of FSA data regardless of source. The data in the warehouse will present a comprehensive, unified collection of FSA data for analytical purposes.

The enterprise data warehouse can enable business groups such as Case Management and Oversight (CMO) to access cross-enterprise data. CMO is responsible for investigating relationships with schools to ensure compliance. Schools are checked for compliance either through a regularly scheduled review, or on an ad hoc basis. The scheduled checks typically use the Postsecondary Education Participants System (PEPS) and eZ-Audit system to gather information. Ad hoc investigations, however, often extract data from many systems such as PEPS, COD, the National Student Loan Data System (NSLDS), and others, depending on what type of compliance issue they are investigating. The CMO group queries these systems to check business rules in order to identify any schools that need further compliance investigation. Schools identified through this process can have an investigative case opened for further inquiry. Caseworkers currently have to go to many systems to investigate an open case. In long-term data strategy vision, the Common Data Architecture will enable the CMO office to access a consolidated picture of FSA business data. The data warehouse solution will enable investigators to find and collect the information needed to investigate a case, independent of the particular source system or back-end process responsible for sourcing the data. This could both reduce the complexity of investigation, and the risk of using "bad" data.

Possible Benefits:

- Consolidated data warehouse can reduce the workload of CMO. Their systems will be able to retrieve data from one source without having to build links to several systems
- Eliminate manual steps – Caseworkers will not be required to manually enter results and findings back into the PEPS system. Depending on the design and implementation,

case-tracking information could potentially live in the same data store, further reducing data replication.

- Enable enterprise-wide visibility – the Common Data Warehouse can enable access to consolidated school and student data stored in a single location.
- As the business intelligence capabilities are developed, the CMO group can shift from performing scheduled queries to identify new cases and could enable a proactive notification from the data warehouse that automatically flags trading partners as soon as they should open a new case.

3.3 Impacts of a Data Warehouse Solution

The data warehouse will be able to provide data to systems that require analytical or historical data. Regardless of the source or type of data needed, the analytical data architecture will be able to provide the data to any requesting system.
System access to data is generally categorized into one of the following methods:

- Predefined, scheduled outputs to down-stream systems – e.g. sending to system X, via EAI, a list of new loans originated in the past 24 hours.
- Query-Response – A system sends a request to the DW for information regarding a specific student. The data warehouse returns a response to the requesting system. A prime example of this type of access would be to support the student status web service.
- Automatic notification – A record or notice is sent to another system when certain conditions are met or pre-defined events occur (characteristic of active data warehousing). The CMO organization could be automatically notified as soon as a school changes a piece of key data and a case should be opened.

Data marts, typically focused on one business process, or functional grouping of business processes, can improve the efficiency of day-to-day operations, as well as enable new capabilities (such as new data mining efforts). When discussing data marts, however, it is important to have a well-defined process in place to centrally manage the creation of data marts. Without such safeguards, many enterprises fall into the trap of creating custom data marts without a valid business case. Each data mart adds cost and complexity to the overall solution. The data mart approval process should ensure that requests for data marts must be based on a valid business case and that the capabilities it would enable cannot be supported through already-in-place data access methodologies. Gartner identifies problems associated with multiple data marts:

"On the surface, individual data marts initially appear to provide better business value proposition because of quicker implementation time, lower cost and direct application support. However, to maintain business value and return on investment (ROI), the application would require modifications to exploit changing business requirements and new opportunities."[5]

[5] © Copyright Gartner, Inc. Source: "Data Mart Consolidation: Strengthening Trend of 2002." March 20, 2002.

FSA data marts would be sourced solely from the data warehouse. Data marts should not bring in new data from other sources. If data from outside the warehouse is entered directly in to the marts, then the FSA data warehouse can lose its status as the single source of analytical information. Data marts are not the "system of record" for their data; the data warehouse retains this distinction as the data marts could be completely recreated, if need be, from the data warehouse.

4 What are the Integration Considerations for the Common Data Architecture?

Data is the fundamental driver of FSA's business services and the organization relies on the accuracy of data to support its mission of providing financial aid to students. The data storage and access techniques that can be deployed have many integration points across the enterprise and throughout the lifecycle of the financial aid process. From application processing to loan servicing and institution eligibility, FSA collects and manages a variety of data. Many elements must supplement the data storage architecture in order to bring FSA's services to its customers and support the processes across the lifecycle. The components include:

- Collecting data from end users
- Presenting data for business decision purposes
- Processing data to support business services
- Moving data within the enterprise
- Populating data into different data stores
- Describing what the data means

4.1 Collecting and Presenting Data

FSA utilizes the Web as one way to collect data. Online applications enable FSA to do business electronically with its trading partners. For example, eZ-Audit allows schools to submit their financial information online to support the institutional eligibility process. Another example is the FAFSA on the Web application, which is responsible for gathering much of the data that FSA uses to support a borrower through the financial aid process. There are specific interactions between the front-end Web architecture and the underlying data architecture that support FSA's services that collect data online. The two architectures should work in tandem with one another to provide an electronic means to collect data and accomplish FSA's Web related business objectives.

In addition to collecting data, the Web can provide a presentation layer and user access to the information stored in the common data architecture. For example, Financial Partners currently use an online business intelligence tool to generate predefined and ad hoc reports. The Web environment needs to integrate with the storage architecture so that users can easily retrieve data. The online reports provide a means to display data that exists in the enterprise data stores to multiple user groups. Online business intelligence tools could be applied more prevalently to user groups across the enterprise to provide broader access to data and enable a more consistent method for reviewing information contained at FSA.

As FSA moves to a more consolidated data storage architecture focused on shared services within the financial aid lifecycle, it will become easier for FSA to support its Web usage business objectives. The Common Data Architecture supports the ability to provide a seamless Web usage experience with access to integrated customer data. With a common data store, FSA's portals could have a standard location to retrieve the appropriate data to populate FSA's

web pages with customer information. This would make the ability to present data more manageable and efficient for the enterprise, while supporting the enterprise's goals of providing a common "look and feel," seamless navigation, and enterprise Web content management. Additionally, the personalization and ad hoc capabilities that accompany flexible online business intelligence tools complement FSA's initiative to provide customizable Web experiences.

Not only are there integration points between the Web usage environment and the data storage architecture but there are also additional considerations for the interplay between the services FSA offers and the data storage configuration that supports them. For example, FSA could utilize Web services to enable data lookup capabilities that can be shared across the enterprise. For Web services to provide lookup capabilities, they must have access to the data that enables this service functionality. The diagram below shows an example of the interrelation between Web services, Web applications and the data storage environment:

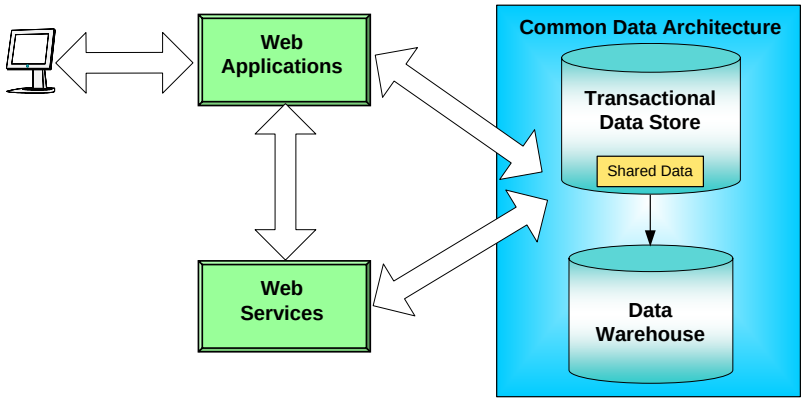


Figure 4.1 - Web Usage and Web Services Integration Example

Web services, like any enabler of system-to-system communication, are only as effective as the quality of the data they transmit. In order for Web services to be an efficient means of providing data access, the data that support their services must be "clean," accurate and accessible. By establishing the Common Data Architecture, FSA can simplify the organization's ability to manage Web services; it would centralize the source that supports the business services. For example, a common data architecture could centrally house a comprehensive view of a borrower's status. A Web service could then expose this functionality for other systems to utilize. This common data storage would allow Web services to operate without having to aggregate data from multiple source systems. Additionally, because the common data architecture would be designed with enterprise business capabilities in mind, the data and status could be more simply and efficiently obtained and uniform information at each phase of the lifecycle could be provided. With the data in disparate systems today, it is possible to obtain status, but the level of detail and supporting information for this status would vary by system. These services could mitigate integration requirements by isolating the warehouse from

individual systems. Gartner recommends: "Where it makes the most sense to do so, shared services should be used to contain the cost of a data warehouse implementation."[6]

4.2 Processing and Moving Data

In relation to the scenarios described in Section 2.0, the evolution from Option C to Option D involves storing data within the shared transactional data store, the data warehouse and source systems outside of the common data architecture platform. In Option C, system specific data could be collected and stored by multiple source system applications. There would also be shared data that exists in a centrally, managed transactional data store. Data from both of these sources may also need to populate the data warehouse. Moving this data from a source system to a central data store could leverage the middleware platform and integrated services functionality discussed in the Internal Data Exchange Strategy (123.1.9). This layer could support the routing and processing demands of the data exchange.

The diagram below illustrates at a high-level the interrelationship of these elements:

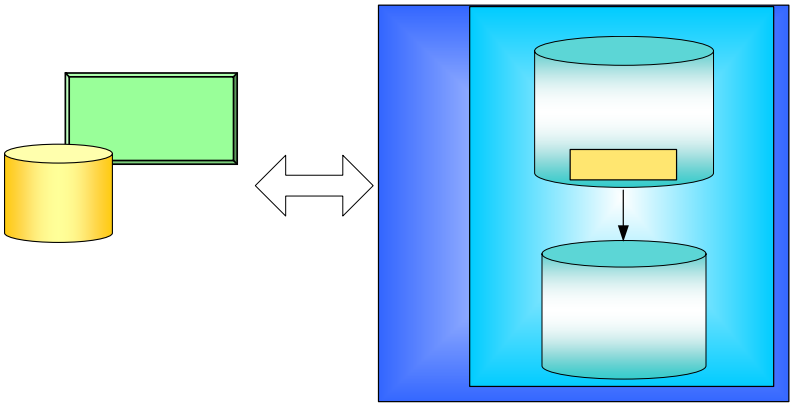


Figure 4.2 - Integrated Services and Data Storage Example

The integrated services functionality as discussed in the Internal Data Strategy (123.1.9) can supplement the data architecture to move data from one point to another and apply business logic, where appropriate, on data that is stored at FSA. Multiple systems might require the same data for system specific processing, such as borrower and eligibility information. The business process management from the integrated services layer could help route data to different systems based on a predefined set of business rules. Within this process, the use of Core Components supports the ability to share data among multiple systems. The Core Components are reusable pieces of business information that provide a standard definition for key data entities across FSA's enterprise. Although the integrated services layer can route and process data, the standardization associated with the use of Core Components alleviates the need for extensive data mapping procedures.

[6] © Copyright Gartner, Inc. Source: "Management Update: Total Value of Opportunity Determination Is Key for Data Warehouse Planning?" September 17, 2003.

The integrated services component of the Internal Data Strategy proposes an environment with loosely coupled systems that enables FSA to bring business services to its customers. This service-oriented architecture supports the data storage transition from Option C to D by allowing the enterprise to move pieces of business functionality through a phased and iterative process. As business capability becomes centrally managed, the data supporting that service can reside in the common data architecture. By creating an environment with centrally managed communication channels and data stores, FSA mitigates the impacts that interface changes will have across the enterprise.

In addition to utilizing the middleware platform and the integrated services layer for data exchange, the common data architecture can leverage population procedures to move data into and within the common data architecture. Population architectures provide uniform processes and services that move and control the movement of data. For example, Extract, Transform and Load (ETL) processes are often used to move large volumes of data. ETL processes can be used when moving data between operational systems and a data warehouse, as well as between data stores within the common data architecture. ETL processes can be custom-built or produced by an ETL tool. In complex environments, both tools and custom code are often utilized.

The majority of ETL processes occur in the realm of data warehousing and decision support (i.e., databases for business analysis and reporting as opposed to transactional or OLTP databases). ETL solutions support the timing considerations required for the purposes of decision support analysis. Data is captured through periodic snapshots and is compiled over time to create a repository of historical data. An ETL process is dedicated to, and optimized for moving large volumes of data to the warehouse in a small window of time. It can be thought of as a huge data pipe that opens to pour a snapshot of data into the warehouse.

The transformational capabilities help ensure that data from different sources are put into a standard format or a format required by the receiving data store. To minimize transformation performed by the ETL process, the data architecture can leverage Core Component definitions for standard and consistent data exchanges. Additionally, data transformation implies that operations may need to be performed on the data. This can include applying business rules and simple and complex calculations. ETL tools can support the need to join, aggregate and lookup data. Finally, ETL processes can copy and transform data from a warehouse into data marts. This allows the data marts to store information in a format that supports the analysis requirements of specific business processes. This includes manipulating data for population into a data mart to allow analytics that might not have been possible, or efficient when attempted against a warehouse.

Overall, the ETL process must be able to interface and efficiently populate the chosen database architecture. The logical and physical distribution of the ETL processes are major components of the population architecture, which supplement the overall data framework and enable the data warehouse to support business decision analysis. The population architecture is an important component in the interface between the integration services, which manage data flow throughout the enterprise and the common data architecture, which houses an enterprise view

of the consolidated data. The diagram below illustrates how the ETL process can fit into the Common Data Architecture:

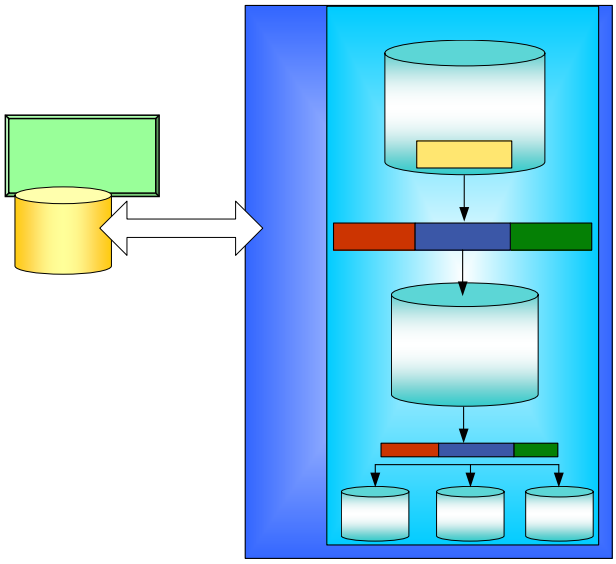


Figure 4.3 - ETL Example

Data from external sources may also be collected by and/or exchanged with FSA. A common data architecture supports FSA's business objective of consolidating its data exchange points with external trading partner systems into a single, virtual entry point called the FSA Gateway. The External Information Access Strategy (123.1.11) further describes this concept in detail. For the purposes of its integration with the common data architecture, data transfers through the FSA Gateway would utilize the integrated services platform for the application of any transformation or business logic before it is handled by the population architecture to finally reside within the common data architecture. For example, a school could submit its direct loan disbursement updates by submitting the Common Record via the FSA Gateway. The integrated services layer could manage the ability for the enterprise to share this data with the systems that need it, such as CSB and FMS. As FSA consolidates its services as proposed by Option D, the integrated services could direct the incoming data to the shared services layer of the common data architecture. In this scenario, the student record would be updated with the new information and the supporting data would update or populate the common operational data store to be used by the lifecycle processes that need it. Utilizing the RID and SSIM common identifiers, the enterprise can efficiently target those records that need updating. The following diagram demonstrates an example of these interrelationships between external data exchange and data storage as it relates to the business scenario described above:

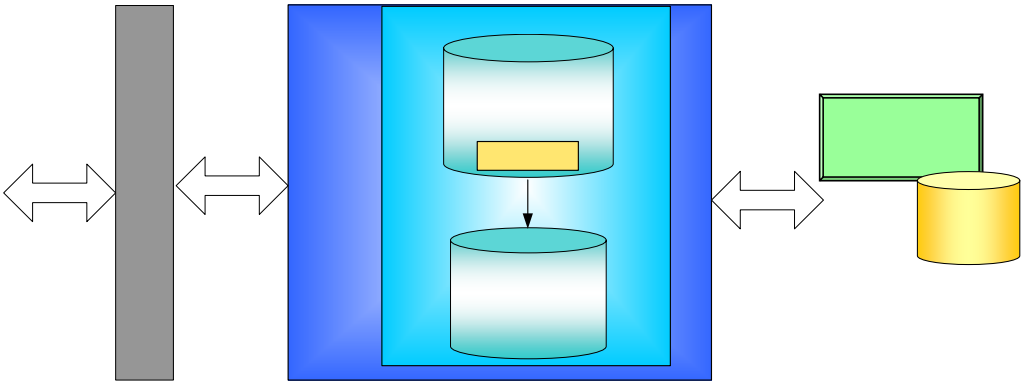


Figure 4.4 - External Information Access and Data Storage

4.3 Describing Data

Information about the definition, use, and/or calculation of individual data fields is a key consideration of any large-scale data integration effort. Metadata is information that lets you catalog, sort and find digital content in different ways. Things like key identifiers, data descriptions, and file types are examples of metadata. If captured and catalogued properly, the metadata provides an information base that can be used to help understand:

- What business information is available.
- The precise meaning of the business information.
- How changes to databases or applications will affect other system components.

Metadata can further be defined as structured data that describes the characteristics of a resource associated with electronic systems. This can be business related (e.g. data dictionary, calculations or process steps) or of a technical nature (e.g. hardware/software information, process statistics, etc.). At a minimum, a consolidated data source needs to be able to provide to its user community with a valid business description for commonly accessed data. An example of this data-element business metadata would be a data dictionary, allowing users to look up the definitions of key fields in the Common Data Architecture.

Without useful descriptions of the data in a data warehouse, business users are confined by their individual knowledge of the data. They are forced to rely on data warehouse support staff, or on a small number of users that hold the knowledge capital associated with the data, in order to navigate to the information they need. The situation actually undermines the entire value proposition of data warehousing, and hinders the opportunity to leverage enterprise information for the benefit of the business processes. It also encourages individual interpretation of the meaning of enterprise data, which leads to incorrect or inconsistent findings and questionable decisions. Gartner identifies that most organizations will find value in a properly managed metadata solution:

Trading Partner
Data Exchange
(ie, Schools)

FSA Gateway

"To maximize success in a service-oriented development of applications and enterprise-wide sharing and reuse initiatives (such as data warehousing), a well-scoped and focused function to manage key metadata artifacts is cost-effective and critical."[7]

Data element definitions may occur throughout the enterprise. The same data element could be categorized and described within an XML repository, or possibly stored in one or many development tools. These varied technology environments might each contain their own, "in-house" definition as to what the element means. As the FSA data storage solution advances from its current environment to its visionary end-state, integration of this and other metadata across platforms should be considered.

Integrating and consolidating metadata into a centrally accessible location allows the enterprise to provide better access to this information for its end users. By providing the data definitions for core operational data and the relevant reference data, an integrated view of metadata could contain the foundational information that could be leveraged by all downstream systems and processes as part of their metadata descriptions.

[7] © Copyright Gartner, Inc. Source: "What is Metadata and Why Should You Care?" April 22, 2003.

5 How will the Data be Stored?

The shift from system-specific storage to a centralized cross-FSA Common Data Architecture is not a simple task. Consolidating data from disparate legacy systems while ensuring all users' access requirements are met is a complex process. FSA should implement the transformation over time, thus achieving incremental benefits. This evolutionary approach will create many decision points relating to how the data storage solution will be implemented. Several considerations must be accounted for as the enterprise evolves from the current state to the target vision.

Data storage decisions are made at one of three levels: conceptual, logical, or physical. These categories are hierarchical in that conceptual questions are the highest-level down to the physical. Conceptual decisions must be made before addressing logical, which must be addressed before physical questions. The technical strategy answers conceptual and logical questions, while providing considerations to begin to address the physical data storage .

A detailed requirements and design analysis is necessary to fully outline the physical architecture necessary. There are considerations that need to be taken into account as FSA migrates from its current state, toward Option C, and finally to the target-state of Option D. Future work toward implementing the Common Data Architecture needs to address each these concerns in detail.

5.1 Data Architecture Considerations

Business owners in the retreats outlined the direction and conceptual end state that FSA will target. This conceptual picture is of value because it provides a vision to understand and define in greater detail. The Techcanical Strategies team, in their subsequent working sessions, built upon the conceptual target state by outlining considerations for the movement from current state to the target state. The following were identified as areas that would require further refinement in order to fully understand the implications and begin to outline a path to achieve the target state:

- Enterprise Data Model
- Database Structure
- Increasing Data Model Complexity
- Conflicting Business Logic

ENTERPRISE DATA MODEL
Data model consolidation should be addressed as system-specific databases are phased out in favor of a centralized data store for FSA front-office support. In today's environment, each system has its own proprietary data model. These data models were developed independently and thus present disparate ways of organizing and storing FSA data. As data in these systems is consolidated into the Common Data Architecture vision, an overall coherent data model must be developed.

DATABASE STRUCTURE

Modern systems focused on transaction processing typically use a normalized relational data model. Normalization is a way of storing data without any redundancy. A borrower's address, for instance, would be stored once in an address entity, rather than as a part of each loan. The main objective of normalized modeling is to reduce this data redundancy for ease of update, insert, and delete performance; it is faster to process an address update in one location, rather than having to find all occurrences of an address and make many separate updates.

Analytical data storage environments are typically in one of two layouts: dimensional or relational. Dimensional data modeling is a system of storing data using facts and dimensions. These facts and dimensions focus on presenting the business relationships between data elements. Relational modeling is a way of storing normalized data through the use of many related tables, each storing a single type of data. Dimensional models are more intuitive than their relational counterparts; end users are better able to understand and use them. Dimensional modeling will produce a star schema. The dimensional approach benefits from shorter, less complicated paths to finding related data than the normalized method. A relational data model can store the data in the smallest possible space, although finding and accessing needed data can require more effort because of the widely dispersed nature of storage. The strengths and weaknesses of these different approaches to database design should be factored into the design of the Common Data Architecture. Components providing operational support should focus more on efficient transaction processing while the data warehouse solution should be optimized for providing analytical and query-able end-user access.

INCREASING DATA MODEL COMPLEXITY

The operational components of the Common Data Architecture vision will need to store data relating to the Awareness, Application, and Delivery lifecycle phases while the data warehouse solution will need to store information on both these front-end, as well as the back-end FSA processes. Each component of the Common Data Architecture will need to have a robust and comprehensive data model to meet these storage requirements. The long-term data storage strategy is the gradual implementation of the Common Data Architecture. Each successive integration step will require the analytical and operational components to store a broader range of data. As the data models grow to accommodate more and more FSA business processes, the complexity of their models will likewise increase. A common risk of large-scale data integration projects is the data model, without an appropriate growth strategy, could grow to become unwieldy. Careful consideration needs to be given to ensure that the complexity of the data model does not increase to the point of making access to either the operational or the analytical components too cumbersome. As complexity of a data model grows, so does the risk of users being unable to effectively use the database(s) to support their business needs. Effective metadata management and end-user access capabilities can mitigate some of the complexity risks by enabling users and systems to better navigate the data warehouse.

CONFLICTING BUSINESS RULES

The consolidation of data storage across business processes presents challenges associated with multiple systems acting upon a single set of data. In the FSA target end-state, front-end data

will be stored in a centralized location, supporting many diverse business processes. In the shared operational database, it will be critical to understand how different processes will update shared data. FSA's current-state has systems applying varied business rules to similar sets of data. When this data is consolidated in the CDA, the overlap and relationships between these business rules will need to be designed to prevent separate processes from applying contradictory logic to the shared data.

5.2 Integration Steps

The long-term FSA vision calls for the incremental implementation from present-state to an end-state Common Data Architecture. Just as this long-term vision should not be attempted in one massive program, nor can it be implemented at too granular a level. The blocks of implementation and integration work need to be large enough to allow for incremental benefit realization without becoming too wide in scope. As data ownership is transferred from today's systems to the operational components of the solution, the data model must be shifted in logical groupings, large enough to support a business process. FSA should not try to move one field at a time, or all systems at once, but rather a logical, manageable grouping of business entities.

5.3 Performance

Performance is always a concern with data integration efforts. Performance considerations must be taken into account for both the operational and the analytical components of the Common Data Architecture. The operational requirements dictate that the data stores manage transactional updates fast enough to support FSAs front-end business processes. Performance of these transactional updates must be closely managed as the size, scope, and complexity of the data model increase. The analytical component of the solution has a different set of performance concerns. The primary focus of the data warehouse is to support the viewing of data. Regardless of the access method used, the data warehouse clients must be able to find the data they need in a reasonable time. The analytical requirements are different in that they are much more focused on logical extraction capabilities, in contrast to supporting transactional updates.

Several factors influence database performance:

- Database Hardware – What type of machine will be used to physically store the data? Each hardware platform presents a unique set of performance advantages and disadvantages. Some hardware solutions work better with certain software packages. The hardware and software solutions cannot be selected independently, as there are many interrelationships between the two.
- Database Management Software – There are several database software platforms on the market. These packages provide varying levels of performance. A key factor for FSA will be how the software package will scale to larger sizes. The iterative nature of the data strategy will require the software solution to support performance requirements for a smaller initial implementation and grow to support a much larger solution at the end-state.

- Fault Tolerance – There are several ways to store data to provide fault tolerance using planned data redundancy. These techniques provide varying levels of backup against physical and software failures. Typically, the more fault tolerant a solution is, the more processing time and disk space is required for each transaction. This is a trade-off that needs to be closely examined with FSA's internal and external (governance) requirements in mind.
- User-base Scalability – The number of users and systems accessing the Common Data Architecture will grow as more systems are integrated, and more end-users are provided access to the system. The ability of the hardware and software packages to scale to larger user bases will be an important factor in choosing all of the components in the Common Data Architecture.

5.4 Stewardship of Data

There are different retention requirements for an operational versus an analytical data store. In the FSA data framework strategy, data would be stored in the student and school/trading partner master for a shorter time, while the data warehouse would have a longer retention (out to several years). These requirements arise from the differing nature of the two data stores. The operational components' primary function is to enable operational processes (daily processing to enable business functions between systems). These processes rarely require "old" historical data. This historical data is most often used for analysis purposes. Minimizing the scope of data in the operational stores will enable better operational performance, while still enabling historical analysis to occur in the data warehouse.

It is important to note that as data "rolls-off" the operational store, it will continue to live solely in the data warehouse. Historical data, therefore, has different systems of record depending on where in the lifecycle it is. The operational data will be the source of record of a borrower's name, for instance, until after loan repayment. Once activity on a loan or borrower has ceased for a certain period of time, that data would then be automatically be removed from the operational system. This data cannot be completely removed, however, due to governance and reporting requirements. From that point on, the data warehouse would "own" the record. A separate archival process will govern the final removal of data from the data warehouse (after a *much* longer time interval.)

DISTRIBUTED VERSUS CONSOLIDATED DATA WAREHOUSING

Should a data warehouse be a set of "federated" data marts, or a centralized database? This question has been the subject of debate in the business intelligence community for quite some time. Current business trends are moving away from the distributed approach and focusing on centralized solutions. The cost of managing a distributed data warehouse environment has become increasingly prohibitive. Data model changes, cross-population between the marts, and access control management across the environment can become increasingly problematic with each additional data mart used in a distributed environment.

A recent Gartner industry analysis classified federated data wareshousing as a solution used "usually as an interim solution to a data mart consolidation strategy."[8]

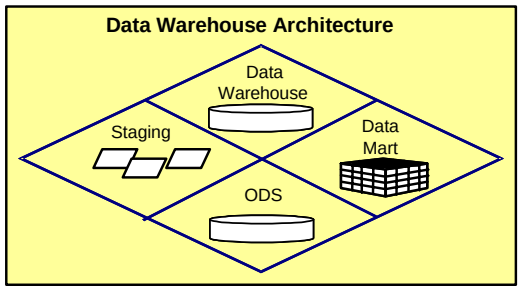
5.5 Technology Support for Data Access and Storage

The physical aspect of a data warehouse solution can be broken-down into its component parts:

- Data Warehouse Architecture
- End-User Access Applications
- Population Architecture
- Metadata Management
- Operations Architecture
- Data Sources

Each functional component can be comprised of multiple technologies. It is important to understand the wide breadth of these technologies available. Once the FSA high-level requirements are fully defined, individual tools should be analyzed for their fit as part of the FSA solution.

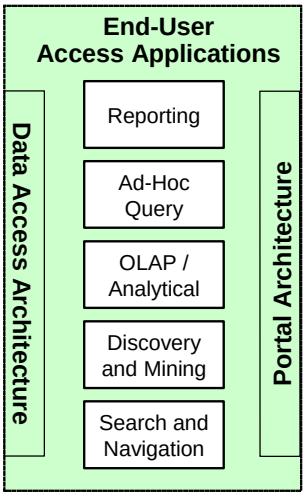
Data Warehouse Architecture



- Relational Data Base Management System (RDBMS) – The RDBMS is the software engine responsible for managing the actual storage and access to the data. FSA should carefully consider scalability requirements when selecting an RDBMS provider.
- Hardware – Physical machine components of the solution including the physical servers, disks, and networking components of the long-term solution
- Software – Includes both operating system and application software packages. Operating Systems (OS's) are used to manage a physical computer, while application packages are the individual software suites that provide specific capabilities.
- Design tools – Primarily modeling suites used to define and manage the FSA warehouse data model. FSA's current design tools should be assessed for fit with target-state environment in the design phase.
- In-place FSA architecture will need to be supplemented in order to accommodate the full CDA requirements due to the massive storage needs and the shift to a data environment shared by many diverse business processes.

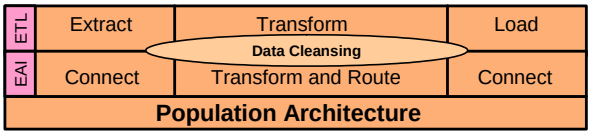
[8] © Copyright Gartner, Inc. Source: "Hype Cycle for Data Warehousing" May 30, 2003

End-User Access



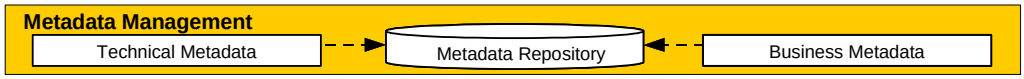
- Reporting/Query – Part of the overall business intelligence capabilities, responsible for providing analytical access to the data warehouse to the FSA community.
- On-Line Analytical Processing (OLAP) – A system for managing a data storage optimized for providing analytical and research access to the data – primarily used for data warehouses and data marts.
- Portal – Integration with the FSA portal is one way of letting end-users access the data warehouse without having to implement or support dedicated client software.
- Data Mining – Non-standard database trend analysis used to allow a specialized group of end users "dig" through the data to identify previously unknown trends and relationships.
- The FSA end-user access architecture needs to be supplemented in order to support the target vision. The user access components in-place today are not deployed across the entire organization, nor are they exposed to enterprise-wide data. In the FSA target vision, end-user access needs to be integrated with Enrollment and Access Management initiatives to standardize access across the enterprise.

Population Architecture



- Extract, Transform, and Load (ETL) – ETL is the process of moving data from disparate source systems into the data warehouse and data mart. ETL can be accomplished in many ways, from custom scripting to an enterprise ETL suite. FSA will use ETL primarily to move data from the consolidated operational store into the data warehouse, and from the warehouse to any data marts.
- Data Cleansing – "Scrubbing" data to enforce data quality standards when loading information to a common data store
- FSA currently uses a mixture of custom and packaged ETL solutions. FSA should conduct a review of its data transformation and loading capabilities to determine fit with the target-state architecture.

Metadata Management

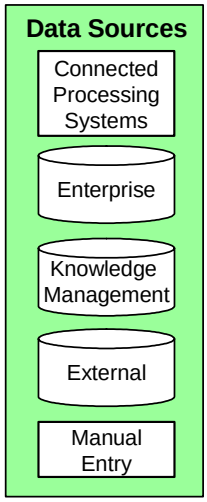


- Metadata – "Data about data", metadata provides users with information about the data they are using. Meta data can also provide a business context to help users ensure they are using the correct data elements.
- Metadata integration can combine warehouse metadata with other information, such as the XML repository, to provide a more comprehensive picture of FSA data.
- FSA needs to build upon its metadata architecture to link or integrate system-specific data dictionaries with enterprise definitions, in order to fully support the Common Data Architecture,

Operations Architecture

- Scheduling – Responsible for managing the load cycles for batch interfaces into the data warehouse. Scheduling is responsible for initiating data population batch jobs and managing any notification and restart/recovery routines.
- Once detailed requirements are defined, FSA should assess its existing operations management solutions to determine if they will be able to support the future requirements of the Common Data Architecture.

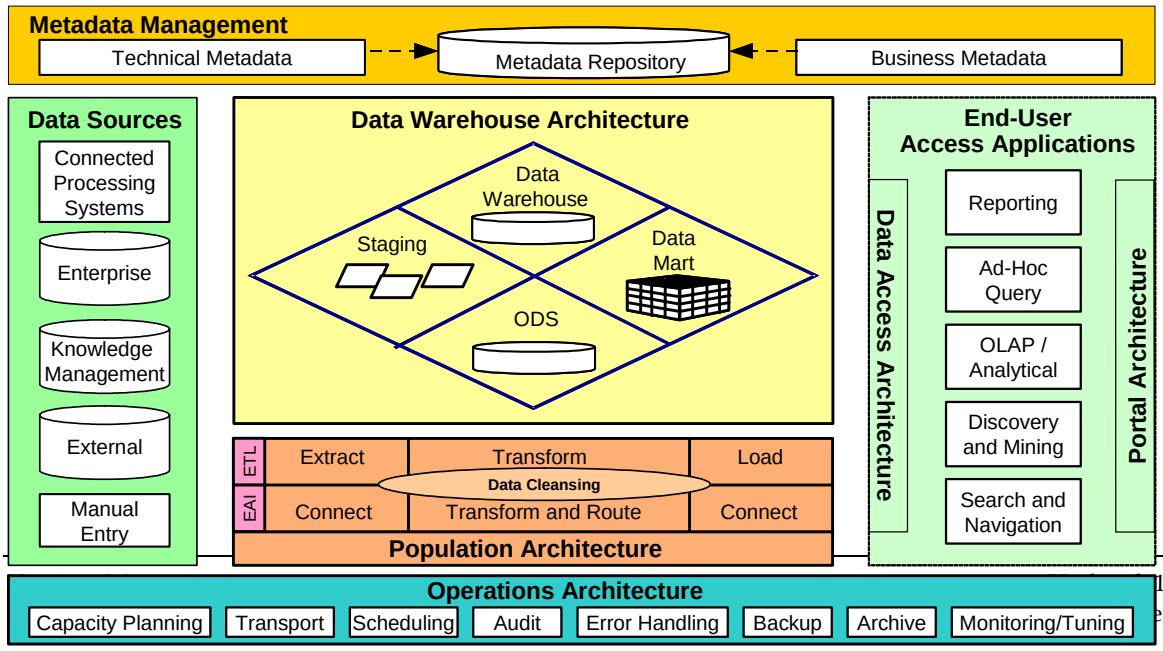
Data Sources



- A data warehouse must be able to communicate effectively with any type of planned source of data. The warehouse must be flexible as these sources often outside the control of an implementation project.
- In the near-term state, FSA will have many systems needing to push data to the warehouse. The mixture of source systems will change as the long-term data vision is implemented. The data warehouse solution needs to accommodate a wide range of planned and unplanned sources.
- Data sources for the Common Data Architecture will be in transition as FSA moves from system-focused to business process-based operations

Composite Data Warehouse Framework

All of these functional components combine to create the overall data warehouse/business integration framework. Successful data warehouse implementations typically contain each of the high-level components. Not every technology listed as part of a component, however, is required. New technologies and capabilities can be added as the scope and use of a data warehouse increases. As FSA moves closer to its target vision, additional tools can be implemented to deliver the full potential business value of an enterprise data warehouse:



6 Recommendation Summary

This recommendation considers FSA business needs, industry trends, and an understanding of the potential return on investment. The rationale for considering a change to the current data architecture at FSA focuses on the following:

- Improving data quality and access to information
- Promoting data consistency
- Driving integration procedures that best align with business processes
- Improving operational efficiencies

This recommendation is only one portion of the overall Data Strategy and it takes into consideration FSA cross-lifecycle business processes. The Data Storage, Management and Access Strategy requires a thorough understanding of all the components that go into the Data Strategy effort in order to realize its full business value. This approach helps to ensure that one component does not negatively impact other areas of FSA. No single solution fits all requirements; the implementation of a new strategy and architecture requires a carefully planned integrated approach.

The ability to provide the right information to the right person in the right format at the right time is a key component of the long-term data strategy. Data storage, management, and access play key roles in enabling that capability. The recommendation to address the high-level business objectives related to this topic is comprised of the following detailed recommendations:

Key Recommendations:

- Store: Develop a Common Data Architecture
  - Build an enterprise data warehouse to store historical information from all phases of the FSA lifecycle.
  - Store shared data to support all front-end business processes.

- Access: Phase-in Uniform Business Intelligence Capabilities Across FSA Organization
  - Create a single point of access for FSA data.
  - Provide tailored access to a broad range of FSA user groups (e.g. reports, ad hoc queries, executive dashboard).
  - Expand centralized web-based access to historical FSA data.
  - Create process-specific data marts sourced from the enterprise data warehouse.

- Integrate: Use Data Storage Solution to Simplify Future Initiatives
  - Enable system and user access to the Common Data Architecture.
  - Leverage internal integration capabilities to support business service requests.
  - Consolidate data definitions (metadata) from multiple systems to provide business context and common definitions to users.
  - Gather feeds from both front- and back-end systems to compile cross-lifecycle information.

- Enable: Develop an Enterprise Data Warehouse Framework
  - Implement robust framework to address all facets of warehouse
  - Build scalable solution to accommodate the increasing scope of shared data and a rising number of students.
  - Long-term solution must accommodate a growing user base as functionality is rolled-out to the broader FSA community.

Implementation of the Data Storage, Management, and Access Strategy could realize the following key benefits:

- Improved Data Visibility
  - Deliver analytical capabilities (detailed reporting to executive dashboards) for business owners to gain visibility across the enterprise.
  - Reduce data gathering effort by tailoring access methods to individual user needs.
  - Improve cross-organizational insight to help understand the state of business and enable quick response to issues that could place FSA on the GAO high-risk list.

- Reduced Maintenance Costs
  - Eliminate redundant data storage; all front-end business processes will act upon the same set of data.
  - Reduce data synchronization processes between systems.

- Reduced Effort Required for Integration Within FSA
  - Simplify new projects by leveraging a consolidated source of data, rather than many independent systems.
  - Leverage security & access management initiatives to provide appropriate user access to data.

- Improved Quality and Usability of FSA Data
  - Reduce data quality issues by storing data once, in a consolidated environment.
  - Optimize performance by separating operational and analytical functions into distinct storage environments.

6.1 Recommendation Details

FSA should implement the following recommendations to fulfill its Data Storage, Management and Access business objectives.

- **Develop a Common Data Architecture**

  The Common Data Architecture is the long-term vision for the FSA Data Framework. As described in Section 2, this recommended approach is for an evolutionary implementation of a centralized data environment. This common architecture will eventually be the single

data store for all of FSA's front-end business processes and enterprise-wide source for analytical data. The details of the Common Data Architecture are being finalized by the Data Framework team and can be found in the Quality Assurance Strategy and Implementation Plan.

Moving from the current-state to the final Common Data Architecture will involve many changes for FSA. The end-state environment will have operational and analytical storage components. The operational elements will be geared toward transaction processing and will support the FSA front-end operational processes. These transactional data storage components will house the Student and School Masters and the Shared Transactional Data, containing the repository of key elements shared amongst the varied business processes. This data will migrate from its current location in the individual source systems, to the shared environment over the course of the implementation. At the end state, this data will live solely in the common architecture. The analytical components of the Common Data Architecture will eventually house a broad range of data from all phases of the aid lifecycle. This analytical storage will be comprised of an enterprise data warehouse and possibly data marts for customized purposes. The data warehouse will form the foundation for FSA's enterprise analytical storage environment. Data access, reporting, and analytical capabilities will be built on top of the warehouse and any associated marts. A key differentiator between this analytical storage and the operational data is that the analytical information is stored in such a way as to enable easier and/or faster data extraction. In contrast to operational systems focus on transaction processing speed, data warehouses are built to empower users to make efficient use of the data for analytical purposes. Rather than transaction speed, analytical storage should be optimized for accessing a wide array of data covering disparate business processes. Like the transactional components, the analytical components should also be phased-in over time. The diagram below outlines the overall evolution:

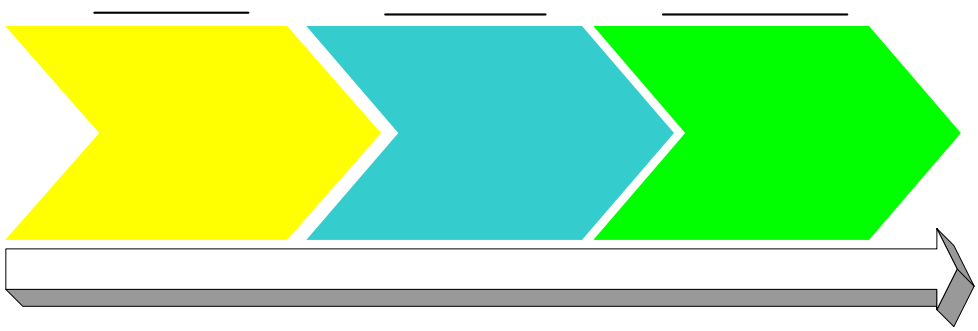


Figure 6.1 - Data Storage, Management and Access Phased Approach

By consolidating all front-end operational data in one place, FSA can reduce or eliminate redundant data storage across the enterprise. Rather than the same piece of data being stored in several operational systems, the data will be stored once, in the common architecture, and will be available to business processes. Transitioning to the target-state

will also enable a reduction in the number of data synchronization processes between systems. Transactional systems will no longer need to go to multiple other systems to obtain all needed information; the transactional components of the Common Data Architecture will act as the single source of data.

The creation of a consolidated source of FSA data will promote improved business insight and enable cross-business process analysis. An example benefit of this capability is the ability to perform verification analysis (the analysis of Pell Grant disbursements with regard to eligible applicants) within a single transactional source, versus matching and reconciling data feeds from two distinct processing systems.

The Common Data Architecture will also enable integrated enterprise views for both students and schools. These centralized sources of data will provide information on the current state of a student or school, regardless of the lifecycle phase. By virtue of data residing in a common architecture, the integrated school and student views will deliver benefits over the current data capabilities by presenting more accurate data. The consolidated storage environment can improve data quality because all front-end processes act upon the data consistently.

- **Phase-in Uniform Business Intelligence Capabilities Across FSA Organization**

Business intelligence (BI) is a system of providing insight into enterprise data in order to support business processes. In order to reap the most benefits from the enterprise data warehouse, FSA should phase-in a single business intelligence platform as the primary access method for end-users. An enterprise-wide BI system will enable users to access data stored in the analytical components of the Common Data Architecture to support many business processes. Most large BI suites currently on the market are able to provide web access to many tools. These tools range from static reports to user-defined, real-time queries. While most of the access methods provided by BI suites can be used through a web front-end, there are typically some specialty applications that require a user to have a local client on their workstation. Such specialty uses typically include data mining and optimizing queries to be shared among many users. Existing BI capabilities exist, but do not run on an enterprise-wide data store. By coupling business intelligence tools with data from the entire organization, FSA can enable a single source of data, capable of reporting on all aspects of the business.

FSA should use a mixture of BI access methods to enable end-users to realize the benefits of a consolidated data warehouse. The business intelligence solution will be the primary access method for end-users to use the data warehouse. Individual BI tools may be reached through the web, by using dedicated client software, or through other communication channels, such as email delivery of reports, or text pager notification. FSA should tailor the available access methods to its varied user groups. Different groups across the FSA community have varied requirements for a business intelligence group. These requirements should be matched against the capabilities and requirements offered by each access method. Several factors need to be considered for each user group and access method including:

- Level of detail – The data warehouse will store data ranging from individual transaction information to high-level aggregates across FSA. At what level on this continuum does each user group need access? Do they need to be able to move up or down in the level of detail?
- Timeliness of data – What are the timing requirements of the user group? For some groups, it may be acceptable to receive a report via email once a month. Other business processes may require visibility into near-real-time intra-day activity. It is important to understand the requirements of the user group in order to provide the right level of data access.
- Training – Some BI access methods are inherently more complex than others. A user receiving a predefined presentation-ready report via email requires much less training than a data quality analyst "digging" for specific technical data within the warehouse. A higher training requirement often accompanies more powerful access capabilities.
- Technical environment – While most BI capabilities are typically available through a web interface, some capabilities may only be accessible with specialized client software on the user's workstation. Distribution and maintenance of dedicated software across a large user base will add cost and complexity to a business intelligence tool. FSA should carefully weigh any special tools maintenance requirements against the size of the user group.

In the past, business intelligence tools have often required such specialized software to enable many, if not all, of the access methodologies. Current industry trends, however, are driving BI providers to enabling web access to their tools, especially for those access methods targeted at the larger user communities. FSA should evaluate the various BI offerings on the market, paying special attention to the types of access provided, and any technical environment requirements on end-users.

Implementation of a single business intelligence suite across the enterprise is a key enabler for providing the right information to the right person in the right format at the right time. The key to this recommendation is distributing access across the entire organization, and possibly to key users outside FSA. The current FSA data marts enable some sense of business intelligence, but the capability is not deployed in a comprehensive manner across the enterprise. With the enterprise BI suite and comprehensive data warehouse, user groups across FSA can have access to organization-wide data, in the format most convenient and efficient for their needs. While the majority of data will come from FSA sources, a robust business intelligence architecture does not preclude bringing in data from other sources. For example, data from other Department of Education systems could be incorporated to support broader analysis capabilities. A modern BI platform will also enable FSA to better manage access to the data to a wide range of users. The business intelligence toolset will be the primary interface for end-users to access data stored in FSA's consolidated analytical warehouse. This consolidated access will enable more efficient management, enhancement, and operation of user access to the warehouse. A recent Gartner analysis of data warehouse costs and benefits highlights the value of a comprehensive approach to business intelligence capabilities:

"Enterprises that do not invest in coordinating and leveraging their business intelligence efforts will spend an order of magnitude more to support multiple, uncoordinated initiatives."[9]

- **Use Data Storage Solution to Simplify Future Initiatives**

Implementation of the Common Data Architecture will work with other initiatives to deliver business capabilities to FSA. Both the consolidation of operational data and the implementation of an enterprise data warehouse will reduce the effort of other integration activities. The common data architecture enables FSA to share services with multiple business processes across the enterprise. Under the proposed model, the integration services layer supports data exchange across the enterprise by executing business logic from a common location, which can aid in the downstream processes of the financial aid lifecycle. By integrating Web applications and Web services capabilities with the data storage solution, FSA can provide more uniform access to enterprise information and standardize the methods of data exchange. This functionality will not only support business decision capabilities but it will also enable the enterprise to operate more efficiently by providing right-time access to data.

The analytical components of the Common Data Architecture will work with the integration services to support FSA business needs and provide better access to financial aid information. The common data architecture can support FSA's need to have access to comprehensive customer information at any point within the financial aid lifecycle and help meet the specific reporting requirements of multiple business entities.

Metadata can provide business benefit by providing information about data. This information, such as a definition of a data element, can enable users to better understand data in the Common Data Architecture. Metadata can be generated and stored in several places and by many processes. FSA should look to integrate this metadata as it moves toward the visionary end-state. One likely candidate for early integration is the metadata stored within the XML repository. The XML repository contains definitions and other data related to many of the data elements currently transmitted between FSA systems. These data elements will eventually be housed in the Common Data Architecture. It would be useful to provide as much meaningful context and definition of data elements to the user community. By providing the end users with metadata contained in both the XML repository as well as any native data definitions in the data warehouse, they will be better able to understand exactly what the data is. Better understanding of the data will help reduce the possibility of misinterpretation of data. Metadata integration should be phased-in as FSA progresses from its current environment, to the Common Data Architecture target solution. A gradual implementation of end-user access to metadata is the recommended approach, in-line with current industry trends:

[9] © Copyright Gartner, Inc. Source: "Management Update: Total Value of Opportunity Determination Is Key for Data Warehouse Planning?" September 17, 2003.

> Even in the long haul, it is unlikely that an enterprise will want to document all its metadata, due to cost and effort vs. return on investment. As a best practice, leading enterprises focus on formally documenting key, mission-critical, widely shared business processes, applications and technologies, and living with less-formal documentation on the metadata in others.[10]

The key to successfully using metadata to support business decisions is to start with a small, manageable group of commonly-used information. Once the initial implementation is in place, FSA should expand its use of metadata over time to support a broader range of user requirements.

Implementing the Data Storage, Management, and Access Strategy should be closely coupled with other Data Strategy initiatives. The creation of a common source for operational and analytical data will enable initiatives such as Web Usage, Web Services, Internal Data, and External Information Access to be implemented with reduced effort. A centralized data store will allow other integration efforts to connect to a single source for operational information. Without consolidated data storage, other projects could possibly have to connect to several source systems to gather their needed information.

- **Develop an Enterprise Data Warehouse Framework**

To be successful, the Data Storage, Management, and Access Strategy must take into account several considerations related to long-term growth throughout its implementation. FSA needs to manage these considerations as it makes decisions regarding the detailed planning and implementation of the data strategy. Because of the evolutionary nature of the FSA data storage strategy, the solution must be able to start small and grow as more business functionality is transferred into the Common Data Architecture.

Key Considerations:

- Data Architecture
- Sizing of Integration Steps
- Performance
- Transitioning Ownership of Data to the Data Warehouse

Bridging these considerations is a fundamental question as to the nature of the analytical environment. Two possible architectures for analytical environments are a system of linked data marts and a centralized data warehouse. A federated system of process-specific data marts could easily cause problems as the amount and scope of data increased. This so-called distributed or virtual data warehousing tends to be difficult to adapt to changing data requirements. Because of increased maintenance and support costs associated with each additional data store in the system. The FSA analytical data environment should be comprised of a centralized data warehouse able to easily scale to accommodate growing data scope. This authoritative historical source will be responsible for long-term ownership of FSA data. Additional data marts could be implemented to support specialized reporting

[10] Source Gartner :"Seven 'Architectural Cuts' of Metadata." May 14, 2003.

needs. Data marts should be sourced solely from the enterprise data warehouse and should not grow to include data not found in the centralized store. If sources other than the warehouse are allowed to feed data to the marts, the marts in effect become separate, smaller data warehouses and the benefits of having a single consolidated source of analytical data are diminished.

6.2 *Bridging of Gaps*

The following table illustrates how this recommendation fills the gaps as derived from the overall Data Storage, Management and Access business objectives:

Business Objective Rank	Gap Description	Fulfilled by Recommendation	Explanation
1	No customer end-to-end visibility across the Financial Aid Lifecycle.	✓	A common, centralized data store can provide end-to-end visibility of FSA customers.
	No consistent use of business intelligence tools to access data.	✓	A common set of BI tools could give users access to DW data.
	Minimal user access to create queries and mine data as needed.	✓	Flexible analytical capability provided through business driven data marts.
2	Data feeds occur at varied frequencies and are not in the required order.	✓	Consistent data feeds from operational store to data warehouse and data warehouse to data marts allow for improved data quality across the enterprise.
	Un-reconciled data is often transferred.	✓	Utilizing common data formats and XML Core components reduces the need to reconcile data after transferring.
3	No enterprise-wide view/store of data exists.	✓	Student and School Master views provide access to enterprise-view of data.
	Data mining capabilities are limited.	✓	Proposed data management configuration proposes improved data mining functionality through BI tool sets.
	FSA control of data is impacted by varied third party systems and owners.	✓	Consolidating data into a central location limits third party systems and owners impact on data.

Table 6.1 - Recommendation Gap Bridging

6.3 Alignment with Business Objectives

Fundamentally, technology that improves business processes makes sense. However, it is the realization of the business value that supports change from the status quo. The matrix below reviews each of the previously defined FSA Data Storage, Management and Access business objectives. Although an assessment of the gaps should help ensure that the business objectives are reasonably met, a specific review of the business objectives is outlined below to ensure the true business needs are enabled by the strategy recommendations. Explanations of the assessment criteria can be found in Appendix A: Business Objective Accommodation Criteria.

Rank	Business Objective	Objective Fulfilment	Explanation
1.	Provide data access to varied resource needs, in the formats necessary to provide meaningful business information and provide data mining and analytics tools that allow simple access to data and facilitate trending, forecasting and the necessary information for business decisions.		• A common BI tool set with canned and ad hoc reporting capability allows for simple access to data. • Data warehouse capabilities can support forecasting and trending needs as defined by the business requirements. • Formats for shared data elements managed by XML Core Components.
2.	Improve the timeliness and efficiency of data feeds to accommodate varied audience data needs.		• By migrating FSA systems' services to a data architecture in which systems interact with a common, centralized data store, FSA can achieve various audiences' business requirements related to data access.
3.	Store, exchange, and manage data in a system-independent format that enables access to key data across FSA throughout all phases of the lifecycle.		• Centralized data storage proposes that FSA migrate to a system-independent data architecture that can provide a common means to access data across all phases of the financial aid lifecycle.

Table 6.2 - Recommendation Business Objective Fulfillment

Appendix A: Business Objective Accommodation Criteria

The table below provides a rating scale that explains the measures and criteria used to evaluate how well a particular technical solution satisfies FSA's business objectives. These indicators appear in the Executive Summary and Recommendation sections with respect to both the current and the recommended states for the Data Storage, Management and Access Strategy.

Rating Indicator	Synopsis	Criteria
	Fully Accommodated Objective	The business objective is being completely satisfied.
	Well Accommodated Objective	The business objective is mostly satisfied.
	Partially Accommodated Objective	Some business objective criteria in place, but the business objective is only partially fulfilled.
	Minimally Accommodated Objective	Few parts of the business objective are satisfied.
	Capability Not Accommodated	The solution does not have this element in place and the business objective is not met.

Business Objective Accommodation Criteria

Appendix B: Meeting Minutes – Working Session #1

Date: Monday, 09/08/2003

Time: 10:00 AM – 12:00 Noon

Location: 221A

Objectives:

- Review Data Strategy Business Objectives
- Address key decisions pertaining to the Data Storage, Management, and Access Strategy.
- Review executive retreat direction (including overview of options C & D)

Attendees:

Name	Business/System Area	E-Mail	Phone (Work)	Attendance
Brown, Nate		nathan.r.brown@accenture.com	202.962.0868	In-Person
Greene, Jim		James.Greene@ed.gov	202.377.3560	In-Person
Hardgrave, Terry		terry.hardgrave@pearson.com	202.377.3238	In-Person
Hill, Denise		denise.hill@ed.gov	202.377.3030	In-Person
Merchant, Denise		Denise.Merchant@ed.gov	202.377.3523	In-Person
Michl, Kyle		kyle.a.michl@accenture.com	202.962.0750	In-Person
Prodgers, Al		allen.prodgers@ed.gov	202.377.3276	Phone
Reed, J.R.		j.r.reed@accenture.com	202.962.0646	In-Person
Rigo, Merlina		Merlina.Rigo@ed.gov	202.377.3352	Phone
Winslow, Marty		marty.winslow@eds.com	202.377.3023	Phone
Woods, Terry		Terry.Woods@ed.gov	202.377.3023	In-Person

Action Items from Previous Meeting Minutes
N/A

Discussion Points

Review of data framework continuum (option C -> option D) by N. Brown

Operations costs decrease as you move left to right on continuum

Implementation costs can increase as you move in same direction

Background on how the decisions were made at BIG

Both C and D contain the concept of realigning aid lifecycle to business process steps, rather than being system-focused. The green boxes represent these processes.

Option C

Shared business services

Shared data lives in common architecture

Each business process still has its own data store – Common data stored in the shared architecture

Option D

Similar to Option C, except that the front-end processes do not have individual data stores – all systems use the Common Data Architecture.

FMS & CSB do not change much between option C and option D

Retreat group identified option D as the long-term vision.

A question was raised as to how other systems would be notified that a change had occurred on the shared data store. If and how there should be a change notification are TBD: the overall flow will be dictated by business logic, but the mechanism of data exchange is TBD.

A question was raised regarding the differences of components in the different options. An example would be the CPS system: it is shown in multiple options, but the database would contain less information (smaller scope) as FSA moves to the right on the continuum.

After reviewing the long-term vision options and retreat direction, the working session walked through the key decision matrix

The question was raised: "should FSA have one enterprise source for data, or would you split the data into different instances based on business process (i.e. a consolidated data warehouse vs. BP-focused data marts)?" - Market trends are moving away from the distributed model because of maintainability costs.

Size and complexity issues were raised using one shared data store – the strategy deliverable should include performance considerations.

Database of record – group discussion on what it means to be a database of record (DBOA). The DBOA is the "owner" of data; in FSA long-term vision, the operational component is the DBOA until the data is rolled off the transactional system and onto the analytical one. The data warehouse would then become DBOA.

Important to note that the data warehouse should be analytics-focused and not support transactional processing

Questions were raised as to the use of/need for an enterprise data model and if there was a point where data model complexity could overpower usefulness. The Data Strategy deliverable will include related considerations.

It is important to note that any data marts would be sourced solely from the data warehouse. The data marts will not receive any external feeds, but are instead customized to enable new analytical capabilities, not achievable directly through the data warehouse. It is still TBD as to

whether a data mart could/should contain data *longer* than the data warehouse (but probable answer is "no").

Most modern Business Intelligence suites enable many types of access through a web interface. Types of access possibly needed by FSA include: email pdf reports, direct queries, OLAP drill-down capabilities, ad hoc queries, business language query generation, fax-on-demand*, voice*.

* not identified as a current requirement, but may be future capability

Access control – Distributed access administration would be an ideal solution, allowing varying levels of administrators to manage all users in their hierarchical tree. This would allow management closer to the end user (e.g. a school administrator would decide what people at that school could have access, rather than attempting to centrally manage all access).

Metadata – it would be nice to eventually have some integrated metadata available including both front-end business data (e.g. field definitions) and "state" data (e.g. when was the last time this field was updated, and by whom?).

Users and suppliers of data need more timely notification that new data has been "accepted" to the system.

The data warehouse and data marts should use the same population platform to load data into their respective systems.

New Action Items

Action Item	Owner	Date Due

Next Meeting

The working session agreed that there were no outstanding issues that would require a second working session. Many of the high-level questions regarding Data Storage, Management, and Access were answered (or at least a direction was established) in the executive retreat sessions. If a large amount of follow-up questions arise, we may decide that a second working session should be scheduled.